


```
DB 121 HSDSPGRLMESYCECTWRTETTGATGOASSLSGRLLEOKAASCHNSYIVLCIENSFWT 180
OY 181 A 181
DB 181 S 181

RESULT 2
061434 PRELIMINARY: PRT: 1140 AA.
ID 061434
AC 061434:
DT 01-NOV-1996 (TREMBLREL. 01, Created)
DT 01-NOV-1996 (TREMBLREL. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLREL. 19, Last annotation update)
DE COLLAGEN (FRAGMENT).
GN COL15A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Abe N., Muragaki Y., Yoshioke H., Inoue H., Ninomiya Y.;
RT "Identification of a novel collagen chain represented by extensive
RT interruptions in the triple-helical region."
RL Cell. Mol. Biol. Res. 196:576-582(1993).
DR EMBL: D17546; BAA04483.1; -.
DR HSSP: P39061; IKOE.
DR MGD: MGI:88449; COL15A1.
DR InterPro: IPR000087; Collagen.
DR Pfam: PF01391; Collagen; 6.
FT NON_TER 1
SQ SEQUENCE 1140 AA; 115156 MW; 8B0C7E6862B3BDPE CRC64;

Query Match 86.9%; Score 840; DB 11; Length 1140;
Best Local Similarity 85.6%; Pred. No. 2,6e-77;
Matches 155; Conservative 15; Mismatches 11; Indels 0; Gaps 0;

OY 1 HSHRDFQPVHLVAVLNSPISGSGMRGIRGADFOCFQOARAVGLAGTRFRLSSRLDLYSI 60
DB 957 HTHQDFQPVHLVAVLNTPLSPISGSGMRGIRGADFOCFQOARAVGLSGTRFRLSSRLDLYSI 1016
OY 61 VRRADRAAVPIVNLKDELLFSPSEALFSGSEGLPKGARIFFSDGKDVLRHPTWPKSVW 120
DB 1017 VRRADGSPVIVNLKDEVLSPSWDSLFSGSGQLGPGARIFSDGKDVLRHPTWPKSVW 1076
OY 121 HSDPGRRLTESYCECTWRTETTGATGOASSLSGRLLEOKAASCHNSYIVLCIENSFWT 180
DB 1077 HSDSPGRLMESYCECTWRTETTGATGOASSLSGRLLEOKAASCHNSYIVLCIENSFWT 1136
OY 181 A 181
DB 1137 S 1137

RESULT 3
062001 PRELIMINARY: PRT: 1774 AA.
ID 062001
AC 062001:
DT 01-NOV-1996 (TREMBLREL. 01, Created)
DT 01-NOV-1996 (TREMBLREL. 01, Last sequence update)
DT 01-OCT-2001 (TREMBLREL. 18, Last annotation update)
DE PROCOLLAGEN, TYPE XVIII, ALPHA 1 PRECURSOR (XVIII) COLLAGEN
DE (PROCOLLAGEN, TYPE XVIII, ALPHA 1) (ALPHA-1 TYPE XVIII COLLAGEN).
GN COL18A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [11]
RP SEQUENCE FROM N.A.
```

```
RC TISSUE-PRIMARY TAIL CULTURE;
RX MEDLINE=94245707; PubMed=8188673;
RA Rehn M., Hinkka E., Philajantemi T.;
RT "Primary structure of the alpha 1 chain of mouse type XVIII collagen,
RT partial structure of the corresponding gene, and comparison of the
RT alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen
RT chain."
RL J. Biol. Chem. 269:13929-13935(1994).
RN [2]
RP SEQUENCE OF 1-562 FROM N.A.
RX MEDLINE=94240112; PubMed=8183894;
RA Rehn M., Philajantemi T.;
RT "Alpha 1(XVIII), a collagen chain with frequent interruptions in the
RT collagenous sequence, a distinct tissue distribution, and homology
RT with type XV collagen."
RL Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).
RN [3]
RP SEQUENCE OF 1-562 FROM N.A.
RX MEDLINE=95181468; PubMed=7876242;
RA Rehn M., Philajantemi T.;
RT "Identification of three N-terminal ends of type XVIII collagen chains
RT and tissue-specific differences in the expression of the corresponding
RT transcripts. The longest form contains a novel motif homologous to rat
RT and Drosophila fibrinogen proteins."
RL J. Biol. Chem. 270:4705-4711(1995).
DR EMBL: U03715; AAC52903.1; JOINED.
DR EMBL: U03716; AAC52903.1; JOINED.
DR EMBL: U03718; AAC52903.1; JOINED.
DR EMBL: U34607; AAC52903.1; JOINED.
DR EMBL: U34608; AAC52903.1; JOINED.
DR EMBL: U34609; AAC52903.1; JOINED.
DR EMBL: U34610; AAC52903.1; JOINED.
DR EMBL: U34611; AAC52903.1; JOINED.
DR EMBL: U34612; AAC52903.1; JOINED.
DR EMBL: U34613; AAC52903.1; JOINED.
DR EMBL: U11637; AAC52179.1; -.
DR HSSP: P39061; IKOE.
DR MGD: MGI:88451; COL18A1.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR000024; Fz.domain.
DR InterPro: IPR001791; Lamlnln_G.
DR InterPro: IPR003129; TSPN.
DR Pfam: PF01392; Fz; 1.
DR Pfam: PF01391; Collagen; 6.
DR SMART: SM00210; TSPN; 1.
DR SMART: SM00603; FRI; 1.
DR SMART: SM00282; LamG; 1.
DR SMART: SM00210; TSPN; 1.
DR PROSITE: PS50038; Fz; 1.
KW Signal.
SQ SEQUENCE 1774 AA; 182229 MW; CF4D9BC9E88EF232 CRC64;

Query Match 86.9%; Score 840; DB 11; Length 1774;
Best Local Similarity 85.6%; Pred. No. 4,7e-77;
Matches 155; Conservative 15; Mismatches 11; Indels 0; Gaps 0;

OY 1 HSHRDFQPVHLVAVLNSPISGSGMRGIRGADFOCFQOARAVGLAGTRFRLSSRLDLYSI 60
DB 1591 HTHQDFQPVHLVAVLNTPLSPISGSGMRGIRGADFOCFQOARAVGLSGTRFRLSSRLDLYSI 1650
OY 61 VRRADRAAVPIVNLKDELLFSPSEALFSGSEGLPKGARIFFSDGKDVLRHPTWPKSVW 120
DB 1651 VRRADGSPVIVNLKDEVLSPSWDSLFSGSGQLGPGARIFSDGKDVLRHPTWPKSVW 1710
OY 121 HSDPGRRLTESYCECTWRTETTGATGOASSLSGRLLEOKAASCHNSYIVLCIENSFWT 180
DB 1711 HSDSPGRLMESYCECTWRTETTGATGOASSLSGRLLEOKAASCHNSYIVLCIENSFWT 1770
OY 181 A 181
DB 1771 S 1771
```

RESULT	4			
090ZD2				
ID	090ZD2	PRELIMINARY;	PRT;	226 AA.
AC	090ZD2;			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	COLLAGEN XVIII (FRAGMENT).			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus			
OX	NCBI_Taxid=10116;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=SPRAGUE-DAWLEY;			
RX	MEDLINE=20227226; Pubmed=10766159;			
RA	Perletti G., Concardi P., Giardin R., Marras E., Piccinini F.,			
RA	Folkman J., Chen L.;			
RT	"Antitumor activity of endostatin against carcinogen-induced rat			
RT	primary mammary tumors."			
RL	Cancer Res. 60:1793-1796(2000).			
DR	EMBL; AF189709; AAF00975.1; '-'			
DR	HSSP; P39061; IJOE.			
FT	NON_TER	1	1	
SO	SEQUENCE	226 AA;	23550 MW;	38B83C0486C0E949 CRC64;

Query Match	84.8%	Score 820	DB 11	Length 226
Best Local Similarity	84.0%	Pred. No. 3.5e-76		
Matches 152	Conservative 14	Mismatches 15	Indels 0	Gaps 0

```

OY      1  HSHRPFQVHLVLVALNSPLSGMGIGIGADFOCCQARAVGLAGTFPAFLSSRLQDLYSI  60
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      43  HTHQDFHVLHLVALNPLPSGGMGINGADFOCCQARAVGLAGTFPAFLSSRLQDLYSI  102
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY      61  VRADRAAVPIVLNKLDELFPSEALFSGSEGPLKPGARLTFSEDKVYLHNPMPQKSV  120
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      103  VRRADRSSVPIVLNKLDELFPSEALFSGSEGPLKPGARLTFSEDKVYLHNPMPQKSV  162
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY      121  HGSDPNGRRLLTESYCETWRTAPATQOASSLLAGRLGGSAAASCHAIYVLVLCEIENSFMT  180
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      163  HGSDPNGRRLLTESYCETWRTAPATQOASSLLAGRLGGRKAESCHANSYIVLVLCEIENSFMT  222
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY      181  A 181
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      223  : 223
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT	5			
09WMU5				
ID	Q9WMU5	PRELIMINARY;	PTT;	171 AA.
AC	Q9WMU5;			
DT	01-NOV-1999 (TrEMBLrel. 12, Created)			
DT	01-NOV-1999 (TrEMBLrel. 12, last sequence update)			
DT	01-DEC-2001 (TrEMBLrel. 19, last annotation update)			
DE	COLLAGEN TYPE XVIII, ALPHA (1) CHAIN (FRAGMENT).			
GN	COL18A1.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID-10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=LIVER.			
RA	Jia J.D., Bauer M., Eberspaecher U., Donner P., Schuppan D.;			
RT	"Partial 3'-terminal cDNA sequence of rat collagen XVIII/endostratin."			
RL	Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=LIVER.			
RA	Jia J.D., Bauer M., Sedlacek N., Ruehl M., Riecken E.O., Schuppan D.;			
RT	"Temporopatial expression of collagen XVIII/endostratin in acute and			

RT chronic liver injuries*;
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases
 DR EMBL: AJ236873; CAB44263.1; -
 DR HSSP: P39061; 1KOE; -
 FT NON-TER 1
 FT NON-TER 1
 FT NON-TER 1
 SO SEQUENCE 171 AA; 18933 MW; 81BE2EE3FC2C8E72 CRC64;

Query Match	80.4%;	Score 777;	DB 11;	Length 171;
Best Local Similarity	85.4%;	Pred. No. 6.3e-72;		
Matches 146;	Conservative 11;	Mismatches 14;	Indels 0;	Gaps 0;

QY 9 VLHVALNLSPLSGGRRGIRGADFOCFQOARVAGIAGFEAFSLSRDLDIVSIYRRADRA 68
 Db 1 VLHVALNTPLSGGRRGIRGADFOCFQOARVAGISGTFRAFLLSRDLDSIYRRADRSS 60
 QY 69 VPIVAKDELFPSPMEALFSGSEGPILKRGARIFSPDGDVLRHPWPQKSWMHGSDNGR 128
 Db 61 VPIVAKDEVLSPSHDITLFSGSGQQLHSGARITFFDGRDVLRRHPAWQKSWMHGSDPSGR 120
 QY 129 RLTESECTWRTTEAPASATGQASSLIGRLRLQGSAAASCHAHVATCIENSEFM 179
 Db 121 RLMESECTWRTTEATGVYTGQASSLSGRLLTEQKAEBSCHNSTIYVATCIENSEFM 171

RESULT 6

ID	093419	PRELIMINARY;	PRT;	1344	AA.
----	--------	--------------	------	------	-----

DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-OCT-2001 (TREMBLrel. 18, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE COLLAGEN XVIII PRECURSOR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxId=9031.
RN [1]
RP
RP MEDLINE=98411346; Pubmed=9738008;
RA Halfter W., Dong S., Schurrer B., Cole G.J.;
RT "Collagen XVIII is a basement membrane heparan sulfate proteoglycan.";
RL J. Biol. Chem. 273:23404-23412(1998).

RP SEQUENCE FROM N.A.
RA Halfeter W., Dong S.;
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases
DR EMBL: AF083440; AAC32294.2; -
DR HSSP: P39061; IKOE.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001791; LamtnIn_G.
DR InterPro: IPR003129; TSPN.
DR Pfam: PF01391; Collagen; 7.
DR Pfam: PF02210; TSPN; 1.
DR SMART: SM00282; LamG; 1.
DR SMART: SM00210; TSPN; 1.
KW signal.
FT SIGNAL 1 26 POTENTIAL.
SQ SEQUENCE 1344 AA; 137402 MW; 7AA366AEF940CCD CRC64

Query Match	79.7%;	Score 771;	DB 13;	Length 1344;
Best Local Similarity	77.6%;	Pred. No. 3.9e-70;		
Matches 142;	Conservative 18;	Mismatches 23;	Indels 0;	Gaps 0;

QY 1 HSHDDQPVHLVALNSENPSLGGMRGIRGADFCQEQARAVGLAGTFRATLSRLDDYSI 60
I - - - - -
Db 1161 HVHDDQPVHLVALNPSLGGMRGIRGADFCQEQARAVGLAGTFRATLSRLDDYSI 1220
I - - - - -
QY 61 VRADRAAVPIVNLKDELLPPSMELFSSGSEGLKRGARITFSFDCKDVLRIHTTPQKSVW 120
I - - - - -

```

DB 1221 VRRADTFAPVYNLRKDEVLFSNMEALFTGSEAPLRAGARILSFDGDIQDSAMPQKSIW 1280
OY 1221 HGSDFPGRRLTESYCEWTREAPSATGQASLLGRLLGQASACHHAYIVLCIENSFMT 180
DB 1281 HGSDFPGRRLTESYCEWTREAPSATGQASLLGRLLGQASACHHAYIVLCIENSFMT 1340
OY 181 ASK 183
DB 1341 AAK 1343

RESULT 7
OQCRT2 PRELIMINARY; PRT; 160 AA.
ID OQCRT2;
AC OQCRT2;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PROCOLLAGEN, TYPE XVIII, ALPHA 1 (FRAGMENT).
GN COL18A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRIONIC HEAD;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shimagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiba H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boftelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Guenichsch S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Suzuki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,
RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlsuk S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK014292; BAB29249.1; -.
DR HSSP: P39061; IKOE.
DR MGD: MGI:88451; Col18a1.
DR NON_TER 1
SQ SEQUENCE 160 AA; 1725 MW; 60F853D77C375D2 CRC64;

```

```

Query Match 74.7%; Score 722; DB 11; Length 160;
Best Local Similarity 85.4%; Pred. No. 2, 6e-66;
Matches 134; Conservative 12; Mismatches 11; Indels 0; Gaps 0;

```

```

OY 25 GTRGADFOCFQOARAVYGLGTFRAFLSSRLQDLYSVRRADRAVYVNLKDELPSPSW 84
DB 1 GTRGADFOCFQOARAVYGLGTFRAFLSSRLQDLYSVRRADRAVYVNLKDELPSPSW 60
OY 85 ALFSGSEGLKPGARIFSDGKDVLRHPTWPKSVHGSDFPNRRLTESYCEWTREAPS 144
DB 61 SLFSGSGGLOLPGARIFSDGKDVLRHPTWPKSVHGSDFPNRRLTESYCEWTREAPS 120
OY 145 ATGQASLLGRLLGQASACHHAYIVLCIENSFMTA 181
DB 121 ATGQASLLGRLLGQASACHHAYIVLCIENSFMTS 157
RESULT 8

```

```

O35206
ID O35206 PRELIMINARY; PRT; 1367 AA.
AC O35206;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE TYPE XV COLLAGEN.
GN COL15A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97480713; PubMed=9339358;
RA Hag P.M., Horelli-Kuitunen N., Eklund L., Palotie A.,
RA Pihlajaniemi T.;
RT "Cloning of mouse type XV collagen sequences and mapping of the
RT corresponding gene to 4B1-3. Comparison of mouse and human alpha 1
RT (XV) collagen sequences indicates divergence in the number of small
RT collagenous domains.";
RL Genomics 45:31-41(1997).
DR EMBL: AF011450; AAC53387.1; -.
DR HSSP: P39061; IKOE.
DR MGD: MGI:88449; Col15a1.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001791; Laminin_G.
DR InterPro: IPR003129; TSPN.
DR Pfam: PF02210; TSPN; 4.
DR Pfam: PF02210; TSPN; 1.
DR SMART: SM00282; LamG; 1.
DR SMART: SM00210; TSPN; 1.
KW Collagen.
SQ SEQUENCE 1367 AA; 140525 MW; A483A1254AF3AEB3 CRC64;

```

```

Query Match 56.3%; Score 544; DB 11; Length 1367;
Best Local Similarity 58.2%; Pred. No. 7, 9e-47;
Matches 103; Conservative 29; Mismatches 41; Indels 4; Gaps 1;

```

```

OY 7 OPLHLYVALNSPISGMRGIRGADFOCFQOARAVYGLGTFRAFLSSRLQDLYSVRRADR 66
DB 1194 RPLHLYVALNTPVAGDIR----ADFQCFQOARAGLSTFRFLSHLDLSTVVKAKR 1249
OY 67 AAVPIVNLKDELFPSEWALFSGSEBGLKPGARIFSDGKDVLRHPTWPKSVHGSDFPN 126
DB 1250 FGLPIVNLKGVLEFNWDSIFSGDGQFMTHPIYFSDGRDWDTPQKVVYHGSNNH 1309
OY 127 GRRRLTESYCEWTREAPSATGQASLLGRLLGQASACHHAYIVLCIENSFMTASK 183
DB 1310 GVRLLVYKCEAWRTDMAYTGFAFSLTGKILDKAVSCANRLIVLCIENSFMTDTR 1366

RESULT 9
O3EOD9 PRELIMINARY; PRT; 1367 AA.
ID O3EOD9;
AC O3EOD9;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE TYPE XV COLLAGEN.
GN COL15A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=97480713; PubMed=9339358;
RA Hag P.M., Horelli-Kuitunen N., Eklund L., Palotie A.,
RA Pihlajaniemi T.;
RT "Cloning of mouse type XV collagen sequences and mapping of the

```


DB 1331 GVRVDNVCAMRTADTAVTGLASPLSTGKILDQKAYSCANRLIVLCIENSFMTDAR 1387
RESULT 11
ID 096770 PRELIMINARY; PRT; 102 AA.
AC 096770;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ENDOSTACTIN VARIANT (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Delinger M.H., Trautmann K., Schluesener H.J.;
RT "Endostatin Promotes delayed secondary damage following traumatic
brain injury."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF333247; AAK50626.1; -.
FT NON_TER 1 1
FT SEQUENCE 102 AA; 11147 MW; ECAC47AA6420947D CRC64;

Query Match 54.7%; Score 529; DB 4; Length 102;
Best Local Similarity 98.0%; Pred. No. 9.1e-47;
Matches 100; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 71 IYNLDELFPSEWALFSGSEGLPKPGARIFSPDGKVLRHPTWPKSVWHSQSPNGRR 130
DB 1 IYNLDELFPSEWALFSGSEGLPKPGARIFSPDGKVLRHPTWPKSVWHSQSPNGRR 60
QY 131 TESYCTWRTPEAPSATGQASSLLGRLGQSAASCHNATIVL 172
DB 61 TESYCTWRTPEAPSATGQASSLLGRLGQSAASCHNATIVL 102

RESULT 12
ID 09VS09 PRELIMINARY; PRT; 581 AA.
AC 09VS09;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CG8645. PROTEIN.
GN CG8645.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RC MEDLINE=20196006; PubMed-10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Butris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Danke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durin K.J., Evangelista C.C., Ferraz C., Fierla S., Fleischmann W.,

RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodak A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Haris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin H.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPreston D.,
RA Merkulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Paclet J.M.,
RA Palazolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AE003560; AAF50621.1; -.
DR HSSP: P39061; IKOE.
DR FlyBase: FBgn0035732; CG8645.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR002088; PPTA.
DR Pfam: PF01391; Collagen.3.
DR PROSITE: PS00904; PPTA; UNKNOWN.1.
SQ SEQUENCE 581 AA; 60772 MW; 19EC1E4BCB477FE7 CRC64;

Query Match 44.5%; Score 430.5; DB 5; Length 581;
Best Local Similarity 48.8%; Pred. No. 1.2e-35;
Matches 82; Conservative 29; Mismatches 56; Indels 1; Gaps 1;

QY 8 PVLHLVALNSPLSGMGRIGADFGCQOARAVGLAGTERAEFLSSRLDLYSIVRRADRA 67
DB 373 PVLHLVALNEPSTGDLQIRGADFACYRQGRRRAGLLGTFRKAFSSVYQDLDTIVRRAD- 431
QY 68 AYPVNLKDELFPSEWALFSGSEGLPKPGARIFSPDGKVLRHPTWPKSVWHSQSPNG 127
DB 432 AYPVNLKDELFPSEWALFSGSEGLPKPGARIFSPDGKVLRHPTWPKSVWHSQSPNG 491
QY 128 RLTESYCTWRTPEAPSATGQASSLLGRLGQSAASCHNATIVL 175
DB 492 ERSMDTYCDAMHSGDHLKSPASNLGDKHLLEQKQSCSKILLICVE 539

RESULT 13
ID 017866 PRELIMINARY; PRT; 650 AA.
AC 017866;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CLE-1C PROTEIN.
GN CLE-1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilita; Rhabditidae;
OC Rhabdilitae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21157401; PubMed-11257122;
RA Ackley B.D., Crew J.R., Elamha H., Pihlajanleml T., Kuo C.J.,
RA Kramer J.M.;
RT "The NC1/endostatin domain of Caenorhabditis elegans type XVIII
collagen affects cell migration and axon guidance."
RL J. Cell Biol. 152:1219-1232(2001).
DR EMBL: AF164959; AAD47825.1; -.
DR HSSP: P39061; IKOE.

DR InterPro: IPR000087; Collagen.
DR Pfam: PF01391; Collagen; 2.
SQ SEQUENCE 650 AA; 69596 MW; 6CF29ED9C16B170E CRC64;

Query Match 38.6%; Score 373; DB 5; Length 650;
Best Local Similarity 44.1%; Pred. No. 1.3e-29;
Matches 75; Conservative 26; Mismatches 63; Indels 6; Gaps 4;

QY 9 VLHLVNLSPGSGMRGIRGADFCFOQARAVGLAGTFRAFLSSRLQDLXSYRRADRAA 68
DB 464 VIMHIALSOPFSGNLHGLRGADLCYREARAAGYTTFRAMLSNVODLVRIHVSVD 522
QY 69 VPIVNLKDELFPSEWALFSGSEGPLKPGARIFSFDDKDVLRHPTWPKSVHSGSDPNGR 128
DB 523 TTVVNVAGHLLFPSSWRSFVNGAQ--MNPBAKLFSFDRHVDVLDNSRWDKRVHSGSKDGI 580
QY 129 RLTESYCETWTEAPSATGQASSLGGRLGQSAAS--CHHAYIVLCIEN 176
DB 581 R-ABOYCDGWRRADSSLTSLAGHISNTSIFQSSGSEKCNKLVLCVEN 629

RESULT 14
Q909K6 PRELIMINARY; PRT; 778 AA.

AC Q909K6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CLE-1B PROTEIN.
GN CLE-1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditicoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=21157401; PubMed=11257122;
RA Ackley B.D., Crew J.R., Elamaa H., Pihlajaniemi T., Kuo C.J.,
RA Kramer J.M.;
RT "The NCL/endostatin domain of Caenorhabditis elegans type XVIII
RT collagen affects cell migration and axon guidance."
RL J. Cell Biol. 152:1219-1232(2001).
DR EMBL: AF164959; AAD47824.1; -.
DR HSSP: P39061; IKOE;
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR003129; TSPN.
DR Pfam: PF01391; Collagen; 2.
DR SMART: SM00210; TSPN; 1.
SQ SEQUENCE 778 AA; 83787 MW; 743A7FC8A7FF1ABD CRC64;

Query Match 38.6%; Score 373; DB 5; Length 778;
Best Local Similarity 44.1%; Pred. No. 1.3e-29;
Matches 75; Conservative 26; Mismatches 63; Indels 6; Gaps 4;

QY 9 VLHLVNLSPGSGMRGIRGADFCFOQARAVGLAGTFRAFLSSRLQDLXSYRRADRAA 68
DB 592 VIMHIALSOPFSGNLHGLRGADLCYREARAAGYTTFRAMLSNVODLVRIHVSVD 522
QY 69 VPIVNLKDELFPSEWALFSGSEGPLKPGARIFSFDDKDVLRHPTWPKSVHSGSDPNGR 128
DB 651 TTVVNVAGHLLFPSSWRSFVNGAQ--MNPBAKLFSFDRHVDVLDNSRWDKRVHSGSKDGI 708
QY 129 RLTESYCETWTEAPSATGQASSLGGRLGQSAAS--CHHAYIVLCIEN 176
DB 709 R-ABOYCDGWRRADSSLTSLAGHISNTSIFQSSGSEKCNKLVLCVEN 757

RESULT 15
Q93336 PRELIMINARY; PRT; 864 AA.
ID Q93336;
AC Q93336;

DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE C36B1.1 PROTEIN.
GN C36B1.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditicoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN (1)
RP SEQUENCE FROM N.A.
RA Leonard N.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=9905613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018(1998).
RN [3]
RP SEQUENCE FROM N.A.
RA White S.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z80215; CAB02278.2; -.
DR EMBL: Z81079; CAB02278.2; JOINED.
DR EMBL: Z81079; CAC42298.1; -.
DR EMBL: Z80215; CAC42298.1; JOINED.
DR InterPro: IPR000087; Collagen.
DR Pfam: PF01391; Collagen; 2.
SQ SEQUENCE 864 AA; 93341 MW; 4A8EDB9B960E56F CRC64;

Query Match 38.6%; Score 373; DB 5; Length 864;
Best Local Similarity 44.1%; Pred. No. 1.3e-29;
Matches 75; Conservative 26; Mismatches 63; Indels 6; Gaps 4;

QY 9 VLHLVNLSPGSGMRGIRGADFCFOQARAVGLAGTFRAFLSSRLQDLXSYRRADRAA 68
DB 678 VIMHIALSOPFSGNLHGLRGADLCYREARAAGYTTFRAMLSNVODLVRIHVSVD 736
QY 69 VPIVNLKDELFPSEWALFSGSEGPLKPGARIFSFDDKDVLRHPTWPKSVHSGSDPNGR 128
DB 737 TTVVNVAGHLLFPSSWRSFVNGAQ--MNPBAKLFSFDRHVDVLDNSRWDKRVHSGSKDGI 794
QY 129 RLTESYCETWTEAPSATGQASSLGGRLGQSAAS--CHHAYIVLCIEN 176
DB 795 R-ABOYCDGWRRADSSLTSLAGHISNTSIFQSSGSEKCNKLVLCVEN 843

Search completed: August 12, 2002, 10:14:25
Job time: 132 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 10:12:33 ; Search time 11.91 Seconds

(without alignments)
594.935 Million cell updates/sec

Title: US-10-080-797-1

Perfect score: 967
Sequence: 1 HSHRDFQVHLVALNSPLS.....SCHHAYIVLCIENSEMTASK 183

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	967	100.0	1516	CA1H_HUMAN	P39060 homo sapien
2	840	86.9	1527	CA1H_MOUSE	P39061 mus musculu
3	529	54.7	1388	CA1E_HUMAN	P39059 homo sapien
4	76	7.9	556	GLI_CHICK	P55878 gallus gall
5	75	7.8	271	Y4BG_RHISN	P55374 rhizobium s
6	74	7.7	296	CRK_XENTLA	P87378 xenopus lae
7	74	7.7	780	STRN_RAT	P70483 rattus norv
8	74	7.7	1215	HD46_HUMAN	Q94907 homo sapien
9	73	7.5	512	PPX_ECOLI	P29014 escherichia
10	72.5	7.5	884	YP67_MYCTU	Q50654 mycobacteri
11	72	7.4	304	CRK_HUMAN	P46108 homo sapien
12	72	7.4	304	CRK_MOUSE	O64010 mus musculu
13	72	7.4	304	CRK_RAT	O63768 rattus norv
14	72	7.4	613	HS75_CANAL	P87222 candida alb
15	70.5	7.3	312	NRT1_CHICK	P55806 gallus gall
16	70	7.2	780	STRN_MOUSE	O55106 mus musculu
17	69	7.1	266	CB21_SINAI	P13851 sinapis alb
18	69	7.1	309	THOE_ECOLI	P45476 escherichia
19	69	7.1	326	THOE_PSEME	O00460 pseudomonas
20	69	7.1	364	YM28_MYCTU	Q10512 mycobacteri
21	69	7.1	462	MYCN_MOUSE	P39066 mus musculu
22	69	7.1	1233	NME3_HUMAN	Q14957 homo sapien
23	68.5	7.1	390	PGK_BUCAL	P57525 buchnera ap
24	68.5	7.1	1289	C5AB_BACUD	O45753 bacillus th
25	68	7.0	853	CAR4_HUMAN	Q9Y239 homo sapien
26	68	7.0	999	MERK_HUMAN	Q12866 homo sapien
27	68	7.0	3491	ERY1_SACER	O03131 saccharopol
28	67.5	7.0	145	ANF_RANCA	P18909 rana catesp
29	67.5	7.0	317	GGH_RAT	O62867 rattus norv
30	67.5	7.0	536	CAR3_HUMAN	O9b257 homo sapien
31	67.5	7.0	579	FAAH_MOUSE	O08914 mus musculu
32	67.5	7.0	594	MG11_MOUSE	O60710 mus musculu
33	67.5	7.0	628	PRPE_ECOLI	P74955 escherichia

34	67.5	7.0	692	1	GYRB_BARBA	P94281 bartonella
35	67.5	7.0	1367	1	VGR2_MOUSE	P35918 mus musculu
36	67.5	7.0	3206	1	POLG_PSBMV	P29152 p genome po
37	67	6.9	250	1	HXB9_HUMAN	P17482 homo sapien
38	67	6.9	577	1	BAG3_MOUSE	Q91471 mus musculu
39	67	6.9	956	1	RRO_PSBMV	P21405 southern be
40	66.5	6.9	368	1	GALT_HUMAN	O60795 homo sapien
41	66.5	6.9	449	1	DHE2_CLOSY	P24295 clostridium
42	66.5	6.9	683	1	ACSI_KILUA	O60011 kluyveromyc
43	66.5	6.9	1548	1	SMCY_MOUSE	O62240 mus musculu
44	66	6.8	250	1	HXB9_MOUSE	P20615 mus musculu
45	66	6.8	295	1	MURB_RICCN	Q921t8 rickettsia

ALIGNMENTS

```

RESULT 1
ID      CA1H_HUMAN          STANDARD:          PRT: 1516 AA.
AC      P39060: 09Y608: 09Y607: 09UK38;
DT      01-FEB-1995 (Rel. 31, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      01-MAR-2002 (Rel. 41, Last annotation update)
DE      Collagen alpha 1(XVII) chain precursor [contains: Endostatin].
GN      COL18A1.
OS      Homo sapiens (human).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OX      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=98164096; PubMed=9503365;
RA      Sareela J., Ylikarppa R., Rehn M., Purmonen S., Pihlajaniemi T.;
RT      Complete primary structure of two variant forms of human type XVIII
RT      collagen and tissue-specific differences in the expression of the
RT      corresponding transcripts.
RL      Matrix Biol. 16:319-328(1996).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=20289799; PubMed=10830953;
RA      Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA      Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA      Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA      Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA      Reichwald A., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA      Rosenthal K., Kudon J., Shibuya K., Kawasaki K., Asakawa S.,
RA      Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA      Minoshima S., Shimizu N., Nordstieck G., Hornischer K., Brandt P.,
RA      Schafie M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA      Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA      Weinmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA      Lehrach H., Reinhardt R., Yaspo M.-L.;
RT      "The DNA sequence of human chromosome 21.";
RL      Nature 405:311-319(2000).
RN      [3]
RP      SEQUENCE OF 834-1516 FROM N.A.
RX      MEDLINE=94245237; PubMed=8188291;
RA      Oh S.P., Warman M.L., Seidlin M.F., Cheng S., Knoll J.H., Timmons S.,
RA      Olsen B.R.;
RT      "Cloning of cDNA and genomic DNA encoding human type XVIII collagen
RT      and localization of the alpha 1(XVIII) collagen gene to mouse
RT      chromosome 10 and human chromosome 21.";
RL      Genomics 19:494-499(1994).
RN      [4]
RP      SEQUENCE OF 1334-1516 FROM N.A.
RX      TISSUE=Placenta;
RA      Zhi-Fong H., Biao L., Wei-Jie Z., Xiang-Fu W.;
RT      "Cloning and expression of human endostatin gene in Escherichia
RT      coli.";
RL      Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN      [5]
RP      INVOLVEMENT IN KNOBLOCH SYNDROME.

```

RA MEDLINE-20400145; PubMed-10942434;
RA Sertie A.L., Sossel V., Camargo A.A., Zatz M., Brahe C.,
RA Passos-Bueno M.R.;
RT "Collagen XVII, containing an endogenous inhibitor of angiogenesis
RT and tumor growth, plays a critical role in the maintenance of retinal
RT structure and in neural tube closure.";
RL Hum. Mol. Genet. 9:2051-2058(2000).
RN [6]
RP VARIANT ASN-1437.
RX PubMed-11606364;
RA Iuganelli P., Suzuki O., Godol P.H., Alves V.A., Sertie A.L.,
RA Zorick T., Soares F., Camargo A., Moreira E.S., di Loreto C.,
RA Moreira-Filho C.A., Simpson A., Oliva G., Passos-Bueno M.R.;
RT "A polymorphism in endostatin, an angiogenesis inhibitor, predisposes
RT for the development of prostatic adenocarcinoma.";
RL Cancer Res. 61:7375-7378(2001).
CC -1- FUNCTION: COLA18A PROBABLY PLAYS A MAJOR ROLE IN DETERMINING THE
CC RETINAL STRUCTURE AS WELL AS IN THE CLOSURE OF THE NEURAL TUBE.
CC -1- FUNCTION: ENDOSTATIN POTENTLY INHIBITS ENDOTHELIAL CELL
CC PROLIFERATION AND ANGIOGENESIS. MAY INHIBIT ANGIOGENESIS BY
CC BINDING TO THE HEPARAN SULPHATE PROTEOGLYCANS INVOLVED IN GROWTH
CC FACTOR SIGNALING.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A SHORT FORM/NC1-303 AND A LONG
CC FORM/NC-483 (SHOWN HERE). ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: PRESENT IN MULTIPLE ORGANS WITH HIGHEST LEVELS
CC IN LIVER, LUNG AND KIDNEY.
CC -1- PTM: POLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- POLYMORPHISM: There is an association between a polymorphism in
CC position 1437 and prostate cancer. Heterozygous Asn-1437
CC individuals have a 2.5 times increased chance of developing
CC prostate cancer as compared with homozygous Asp-1437 individuals.
CC -1- DISEASE: Defects in COL18A1 are a cause of Knobloch syndrome (KS
CC or KNO): an autosomal recessive disorder defined by the occurrence
CC of high myopia, vitreoretinal degeneration with retinal
CC detachment, macular abnormalities and occipital encephalocele.
CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC INTERRUPTED HELICES (FACIT) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF018081: AAC39658.1; -;
DR EMBL: AF018082: AAC39659.1; -;
DR EMBL: AL163302: CAB90482.1; -;
DR EMBL: L22548: AAS1864.1; -;
DR EMBL: AF184060: AAF01310.1; ALT_INIT.
DR HSSP: P39061: 1KOE.
DR GYCSuiteDB: P39060: -.
DR MIM: 120328: -.
DR MIM: 267730: -.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001791; Laminin_G.
DR InterPro: IPR003129; TSPN.
DR Pfam: PF01391; Collagen; 6.
DR Pfam: PF02210; TSPN; 1.
DR SMART: SM00282; LamG; 1.
DR SMART: SM00210; TSPN; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing;
KW Polymorphism.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 1516 COLLAGEN ALPHA 1(XVII) CHAIN.
FT CHAIN 1334 1516 ENDOSTATIN.
FT DOMAIN 24 516 NONHELICAL REGION 1 (NC1).
FT DOMAIN 517 550 TRIPLE-HELICAL REGION 1 (COL1).
FT DOMAIN 551 560 NONHELICAL REGION 2 (NC2).
FT DOMAIN 561 640 TRIPLE-HELICAL REGION 2 (COL2).

FT DOMAIN 641 664 NONHELICAL REGION 3 (NC3).
FT DOMAIN 665 786 TRIPLE-HELICAL REGION 3 (COL3).
FT DOMAIN 787 809 NONHELICAL REGION 4 (NC4).
FT DOMAIN 810 892 TRIPLE-HELICAL REGION 4 (COL4).
FT DOMAIN 893 906 NONHELICAL REGION 5 (NC5).
FT DOMAIN 907 948 TRIPLE-HELICAL REGION 5 (COL5).
FT DOMAIN 949 961 NONHELICAL REGION 6 (NC6).
FT DOMAIN 962 1034 TRIPLE-HELICAL REGION 6 (COL6).
FT DOMAIN 1035 1044 NONHELICAL REGION 7 (NC7).
FT DOMAIN 1045 1077 TRIPLE-HELICAL REGION 7 (COL7).
FT DOMAIN 1078 1089 NONHELICAL REGION 8 (NC8).
FT DOMAIN 1090 1111 TRIPLE-HELICAL REGION 8 (COL8).
FT DOMAIN 1112 1118 NONHELICAL REGION 9 (NC9).
FT DOMAIN 1119 1173 TRIPLE-HELICAL REGION 9 (COL9).
FT DOMAIN 1174 1186 NONHELICAL REGION 10 (NC10).
FT DOMAIN 1187 1204 TRIPLE-HELICAL REGION 10 (COL10).
FT DOMAIN 1205 1516 NONHELICAL REGION 11 (NC11).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 129 129 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 164 164 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 691 691 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1329 1329 O-LINKED (GALNAC. . .).
FT DISULFID 1366 1506 BY SIMILARITY.
FT DISULFID 1468 1498 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 1095 1097 MISSING (IN SHORT ISOFORM).
FT VARSPIC 1 180 HTTEAGTLPAPRPSPSGRPAPLPATGSPVPPSS -> MA
FT VARSPIC 181 215 PRCPPMPRRPRLDVLVAPIVLLGVRAASAPR (IN
FT SHORT ISOFORM).
FT VARIANT 1437 1437 D -> N.
FT 428 /FTID=VAR_012709.
FT 841 F -> S (IN REF. 2).
FT 841 I -> V (IN REF. 2).
FT 877 V -> L (IN REF. 3).
FT 886 P -> R (IN REF. 3).
FT 912 P -> R (IN REF. 3).
FT 933 R -> L (IN REF. 3).
FT 975 P -> L (IN REF. 3).
FT 1064 A -> P (IN REF. 3).
FT 1084 L -> K (IN REF. 3).
FT 1120 P -> A (IN REF. 3).
FT 1123 P -> A (IN REF. 3).
FT 1126 P -> PGP (IN REF. 2).
FT 1206 G -> GO (IN REF. 3).
FT 1304 R -> G (IN REF. 3).
FT 1314 A -> G (IN REF. 3).
FT 1324 IR -> CG (IN REF. 3).
FT 1324 R -> T (IN REF. 4).
FT 1443 S -> Y (IN REF. 4).
FT 1483
SQ SEQUENCE 1516 AA; 153840 MW; 3C70F29A4476EE76 CRC64;

Query Match 100.0%; Score 967; DB 1; Length 1516;
Best Local Similarity 100.0%; Pred. No. 2.8e-86;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HSHRDEQVPLHLVALNPSGSGMRGIRGADFOCARAVAGTAGTFRAFLLSRDLDSI 60
DB 1334 HSHRDEQVPLHLVALNPSGSGMRGIRGADFOCARAVAGTAGTFRAFLLSRDLDSI 1393
OY 61 VRRADRAVPIVNLKDELLFPSEWALEFSGSEGPLKRGARIFSGDKDVLRRPTWQKSYW 120
DB 1394 VRRADRAVPIVNLKDELLFPSEWALEFSGSEGPLKRGARIFSGDKDVLRRPTWQKSYW 1453
OY 121 HGSDEPGRRLTESYCEWTWTAPSATGQASSLLGRLGLQSAASCHHATIVYCIENSFT 180
DB 1454 HGSDEPGRRLTESYCEWTWTAPSATGQASSLLGRLGLQSAASCHHATIVYCIENSFT 1513
OY 181 ASK 183
DB 1514 ASK 1516

```

RESULT 2
CAH MOUSE STANDARD; PRT: 1527 AA.
ID P39061; 062002; 061437;
AC 01-FEB-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 1(XVIII) chain precursor [contains: Endostatin].
GN COL18A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A. (SHORT ISOFORM).
RC STRAIN=BALE/C; TISSUE=Liver.
RX MEDLINE=94245707; PubMed=8188673;
RA Rehman M.V., Hintikka E., Pihlajaniemi T.;
RT Primary structure of the alpha 1 chain of mouse type XVIII collagen,
RT partial structure of the corresponding gene, and comparison of the
RT alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen
RT chain.
RN J. Biol. Chem. 269:13929-13935(1994).
RL [2]
RP SEQUENCE FROM N.A. (SHORT AND LONG ISOFORMS).
RA Rehman M., Hintikka E., Pihlajaniemi T.;
RT Characterization of the mouse gene for the alpha-1 chain of type
RT XVIII collagen (COL18A1) reveals that the three variant N-terminal
RT polypeptide forms are transcribed from two widely separated
RT promoters.
RN Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
RL [3]
RP SEQUENCE OF 213-1140 FROM N.A. (SHORT ISOFORM).
RX MEDLINE=94240112; PubMed=8183894;
RA Rehman M.V., Pihlajaniemi T.;
RT Alpha 1(XVIII), a collagen chain with frequent interruptions in the
RT collagenous sequence, a distinct tissue distribution, and homology
RT with type XV collagen.
RN Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).
RL [4]
RP SEQUENCE OF 240-1527 FROM N.A.
RX TISSUE=Liver;
RX MEDLINE=94240111; PubMed=8183893;
RA Oh S.P., Kamagata Y., Muragaki Y., Timmons S., Ooshima A., Olsen B.R.;
RT Isolation and sequencing of cDNAs for proteins with multiple domains
RT of Gly-Xaa-Yaa repeats identify a distinct family of collagenous
RT proteins.
RN Proc. Natl. Acad. Sci. U.S.A. 91:4229-4233(1994).
RL [5]
RP CHARACTERIZATION OF ENDOSTATIN, AND PARTIAL SEQUENCE.
RX MEDLINE=97160848; PubMed=9008168;
RA O'Reilly M.S., Boehm T., Shing Y., Fukai N., Vastos G., Lane W.S.,
RA Flynn E., Birkhead J.R., Olsen B.R., Folkman J.;
RT Endostatin: an endogenous inhibitor of angiogenesis and tumor
RT growth.
RN Cell 88:277-285(1997).
RL [6]
RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF ENDOSTATIN.
RX MEDLINE=98169382; PubMed=9501087;
RA Hohenester E., Sasaki T., Olsen B.R., Timpl R.;
RT Crystal structure of the angiogenesis inhibitor endostatin at 1.5-A
RT resolution.
RN EMBO J. 17:1656-1664(1998).
RL -I- FUNCTION: ENDOSTATIN POTENTIALLY INHIBITS ENDOTHELIAL CELL
CC PROLIFERATION AND ANGIOGENESIS. MAY INHIBIT ANGIOGENESIS BY
CC BINDING TO THE HEPARAN SULFATE PROTEOGLYCAN INVOLVED IN GROWTH
CC FACTOR SIGNALING.
CC -I- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A SHORT FORM AND A LONG FORM
CC (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -I- PTM: PROLINES ARE HYDROXYLATED IN SOME OR ALL OF THE CHAIN.
CC -I- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH

```

```

CC INTERRUPTED HELICES (FACIT) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L16898; AAA37434.1; -
CC EMBL: U03714; AAA20657.1; -
CC EMBL: U03715; AAC52901.1; -
CC EMBL: U34606; AAC52901.1; JOINED.
CC EMBL: U34608; AAC52901.1; JOINED.
CC EMBL: U34609; AAC52901.1; JOINED.
CC EMBL: U34610; AAC52901.1; JOINED.
CC EMBL: U34611; AAC52901.1; JOINED.
CC EMBL: U34612; AAC52901.1; JOINED.
CC EMBL: U34613; AAC52901.1; JOINED.
CC EMBL: U03716; AAC52901.1; JOINED.
CC EMBL: U03718; AAC52901.1; JOINED.
CC EMBL: U03719; AAC52902.1; -
CC EMBL: U34607; AAC52902.1; JOINED.
CC EMBL: U34608; AAC52902.1; JOINED.
CC EMBL: U34609; AAC52902.1; JOINED.
CC EMBL: U34610; AAC52902.1; JOINED.
CC EMBL: U34611; AAC52902.1; JOINED.
CC EMBL: U34612; AAC52902.1; JOINED.
CC EMBL: U03716; AAC52902.1; JOINED.
CC EMBL: U03718; AAC52902.1; JOINED.
CC EMBL: U11636; AAC52178.1; -
CC EMBL: L22545; AAA19787.1; -
CC PDB: 1KOE; 16-FEB-99.
CC MGD: MGI:88451; COL18A1.
CC InterPro: IPR000087; Collagen.
CC InterPro: IPR001791; Laminin_G.
CC InterPro: IPR003129; TSPN.
CC Pfam: PF01391; Collagen; 8.
CC Pfam: PF02210; TSPN; 2.
CC SMART: SM00282; LamG; 1.
CC SMART: SM00210; TSPN; 1.
CC Extracellular matrix;
CC Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing;
CC 3D-structure.
CC STGNL 1 26
CC CHAIN 27 1527
CC CHAIN 1344 1527
CC DOMAIN 27 538
CC DOMAIN 539 565
CC DOMAIN 566 575
CC DOMAIN 576 649
CC DOMAIN 650 673
CC DOMAIN 674 795
CC DOMAIN 796 818
CC DOMAIN 819 901
CC DOMAIN 902 915
CC DOMAIN 916 957
CC DOMAIN 958 970
CC DOMAIN 971 1043
CC DOMAIN 1044 1053
CC DOMAIN 1054 1086
CC DOMAIN 1087 1098
CC DOMAIN 1099 1122
CC DOMAIN 1123 1129
CC DOMAIN 1130 1181
CC DOMAIN 1182 1194
CC DOMAIN 1195 1212
CC DOMAIN 1213 1527
CC CARBOHYD 338 338
CC CARBOHYD 700 700
CC DISULFID 1376 1516
CC POTENTIAL.
CC COLLAGEN ALPHA 1(XVIII) CHAIN.
CC ENDOSTATIN.
CC NONHELICAL REGION 1 (NC1).
CC TRIPLE-HELICAL REGION 1 (COL1).
CC NONHELICAL REGION 2 (NC2).
CC TRIPLE-HELICAL REGION 2 (COL2).
CC NONHELICAL REGION 3 (NC3).
CC TRIPLE-HELICAL REGION 3 (COL3).
CC NONHELICAL REGION 4 (NC4).
CC TRIPLE-HELICAL REGION 4 (COL4).
CC NONHELICAL REGION 5 (NC5).
CC TRIPLE-HELICAL REGION 5 (COL5).
CC NONHELICAL REGION 6 (NC6).
CC TRIPLE-HELICAL REGION 6 (COL6).
CC NONHELICAL REGION 7 (NC7).
CC TRIPLE-HELICAL REGION 7 (COL7).
CC NONHELICAL REGION 8 (NC8).
CC TRIPLE-HELICAL REGION 8 (COL8).
CC NONHELICAL REGION 9 (NC9).
CC TRIPLE-HELICAL REGION 9 (COL9).
CC NONHELICAL REGION 10 (NC10).
CC TRIPLE-HELICAL REGION 10 (COL10).
CC NONHELICAL REGION 11 (NC11).
CC TRIPLE-HELICAL REGION 11 (COL11).
CC N-LINKED (GLCNAC. . .) (POTENTIAL).
CC N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```
FT DISULFID 1478 1508
FT SITE 1104 1106
FT VARSPIC 1 212
FT VARSPIC 213 238
FT CONFLICT 900 900
FT CONFLICT 947 947
FT CONFLICT 964 964
FT CONFLICT 1157 1157
FT CONFLICT 1266 1266
FT CONFLICT 1276 1276
FT CONFLICT 1437 1437
SQ SEQUENCE 1527 AA; 156008 MW; 9645045AF140B513 CRC64;

Query Match
Best Local Similarity 86.9%; Score 840; DB 1; Length 1527;
Matches 155; Conservative 15; Mismatches 11; Indels 0; Gaps 0;

QY 1 HSHRDFQVPLHLVVALNSPLSGMRGIRGADFOCFQOARAVAGLAGTFRALSSRLQDLYST 60
DB 1344 HTHODEQVPLHLVVALNTPLSGMRGIRGADFOCFQOARAVAGLAGTFRALSSRLQDLYST 1403
QY 61 VRADAAAPVIVLKDLELFPSSWEALFSGSEGLKPGARITFSDGKDYLRHPTWPKSW 120
DB 1404 VRADAGSVPIVVLKDEVLSFSGSQCOLQPGARITFSDGKDYLRHPTWPKSW 1463
QY 121 HSDPNGRRLTESYCEFTWTEAPATGQASLLGRLGOSASHHAVIVLCIENSFMT 180
DB 1464 HSDPNGRRLMESYCEFTWTEATGATGQASLLGRLGKASCHNSYIVLCIENSFMT 1523
QY 181 A 181
DB 1524 S 1524

RESULT 3
CABLE_HUMAN STANDARD; PRT: 1388 AA.
ID CABLE_HUMAN
AC P39059;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 1(XV) chain precursor.
GN COL15A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=umbilical cord;
RX MEDLINE=94148920; PubMed=8106446;
RA Kivirikko S., Helenakki P., Rehn M.V., Honkanen N., Myers J.C.,
RA Pihlajaniemi T.;
RT "Primary structure of the alpha 1 chain of human type XV collagen and
RT exon-intron organization in the 3' region of the corresponding
RT gene.";
RL J. Biol. Chem. 269:4773-4779(1994).
RN [2]
RP SEQUENCE OF 1-569 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=94140817; PubMed=8307960;
RA Murgaki Y., Abe N., Nlomiyi Y., Olsen B.R., Ooshima A.;
RT "The human alpha 1(XV) collagen chain contains a large amino-terminal
RT non-triple helical domain with a tandem repeat structure and homology
RT to alpha 1(XVIII) collagen.";
RL J. Biol. Chem. 269:4042-4046(1994).
RN [3]
RP SEQUENCE OF 544-1252 FROM N.A.
RX MEDLINE=93066196; PubMed=11279671;
RA Myers J.C., Kivirikko S., Gordon M.K., Dion A.S., Pihlajaniemi T.;
RT "Identification of a previously unknown human collagen chain, alpha
```

```
RT 1(XV), characterized by extensive interruptions in the triple-helical
RT region.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:10144-10148(1992).
CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN INTERNAL ORGANS
CC SUCH AS ADRENAL GLAND, PANCREAS AND KIDNEY.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC INTERRUPTED HELICES (FACIT) FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; L25286; AAS8429.1; -.
DR EMBL; D12330; BAA04762.1; -.
DR EMBL; L01697; -; NOT_ANNOTATED_CDS.
DR HSSP; P39061; IKOE.
DR MIM; 120325; -.
DR InterPro; IPR000097; Collagen.
DR InterPro; IPR001791; Lamlnln_G.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01391; Collagen; 4.
DR Pfam; PF02210; TSPN; 1.
DR SMART; SM00282; Lamg; 1.
DR SMART; SM00210; TSPN; 1.
DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Cell adhesion; Collagen; Glycoprotein; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1388
FT FT 26 1388
FT DOMAIN 26 555
FT DOMAIN 56 573
FT DOMAIN 574 618
FT DOMAIN 619 732
FT DOMAIN 733 763
FT DOMAIN 764 798
FT DOMAIN 799 822
FT DOMAIN 823 867
FT DOMAIN 868 878
FT DOMAIN 879 949
FT DOMAIN 950 983
FT DOMAIN 984 1013
FT DOMAIN 1014 1027
FT DOMAIN 1028 1045
FT DOMAIN 1046 1052
FT DOMAIN 1053 1107
FT DOMAIN 1108 1117
FT DOMAIN 1118 1132
FT DOMAIN 1133 1388
FT DOMAIN 1358 555
FT REPEAT 358 408
FT REPEAT 409 459
FT REPEAT 460 509
FT REPEAT 510 555
FT CARBOHYD 306 324
FT CARBOHYD 324 324
FT CARBOHYD 687 687
FT CARBOHYD 807 807
FT CARBOHYD 814 814
FT CARBOHYD 1046 1046
FT CONFLICT 10 10
FT CONFLICT 49 49
FT CONFLICT 95 95
FT CONFLICT 150 150
FT CONFLICT 204 204
FT CONFLICT 409 409
SQ SEQUENCE 1388 AA; 141930 MW; 60822AD925A3093D CRC64;
```



```
DB 260 -KOELEMAETWK 271
::: |||:
RESULT 6
CRK_XENLA STANDARD: PRT: 296 AA.
ID CRK_XENLA
AC P87378:
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE SH2/SH3 adaptor crk (Adaptor molecule crk) (CRK).
GN CRK.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxId=8335;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-97180926; PubMed-9029144;
RA Evans E.K., Lu W., Strum S.L., Mayer B.J., Kornbluth S.;
RT "Crk is required for apoptosis in Xenopus egg extracts.";
RL EMBL J 16:230-241(1997).
CC -1- FUNCTION: REQUIRED FOR APOPTOSIS IN XENOPUS EGG EXTRACTS.
CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 SH3 DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
DR EMBL: U89774; ABA49698.1; -
DR HSRP: 064010; 1CKA.
DR InterPro: IPR000980; SH2.
DR Pfam: PF00017; SH2; 1.
DR Pfam: PF00018; SH3; 1.
DR PRINTS: PR00401; SH2DOMAIN.
DR PRINTS: PR00452; SH3DOMAIN.
DR SMART: SM00324; SH2; 1.
DR SMART: SM00326; SH3; 2.
DR PROSITE: PS50001; SH2; 1.
DR PROSITE: PS50002; SH3; 1.
DR SH2 domain; SH3 domain; Repeat.
FT DOMAIN 13 112 SH2.
FT DOMAIN 125 185 SH3 1.
FT DOMAIN 249 289 SH3 2.
SQ SEQUENCE 296 AA; 33409 MW; 544F11FA1F75A66 CRC64;

Query Match 7.7%; Score 74; DB 1; Length 296;
Best Local Similarity 32.4%; Pred. No. 5.4;
Matches 24; Conservative 12; Mismatches 20; Indels 18; Gaps 5;

DB 136 FIGNDEDEDLPFKK-----DIIRIKREKPEQWMAEDSDGRKGMIPVIVKTR-- 184
OY 87 FGSSEG---PLKPGARIFSGDKVLRHPTWPOKSWHGSDPGR--LTESYCEWRT 141
DB 136 FIGNDEDEDLPFKK-----DIIRIKREKPEQWMAEDSDGRKGMIPVIVKTR-- 184
OY 142 APSATGQASLLGG 155
DB 185 PSSPG--SALIGG 196

RESULT 7
STRN_RAT STANDARD: PRT: 780 AA.
ID STRN_RAT
AC P70483:
DT 01-NOV-1997 (Rel. 35, Created)

DB 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Striatin.
GN STRN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 222-243; 258-269; 451-456; 488-494
RP AND 764-777.
RA STRAIN-WISTAR: TISSUE-Brain;
RX MEDLINE-96354919; PubMed-8769426;
RA Castets F., Bartoli M., Barnier J.V., Baillet G., Salin P.,
RA Moglich A., Bourgois J.-P., Denizot F., Rougon G., Calothy G.,
RA Monnerot A.;
RT "A novel calmodulin-binding protein, belonging to the WD-repeat
RT family, is localized in dendrites of a subset of CNS neurons.";
RL J. Cell Biol. 134:1051-1062(1996).
CC -1- FUNCTION: BINDS CALMODULIN IN A CALCIUM DEPENDENT MANNER. MAY
CC FUNCTION AS SCAFFOLDING OR SIGNALING PROTEIN.
CC -1- SUBUNIT: INTERACTS WITH PROTEIN PHOSPHATASE 2A (PP2A) (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC AND MEMBRANE-BOUND.
CC -1- TISSUE SPECIFICITY: MAINLY EXPRESSED IN THE CENTRAL NERVOUS
CC SYSTEM. MOSTLY CONFINED IN DENDRITES, NOT IN AXONS, AND IS MOST
CC ABUNDANT IN DENDRITIC SPINES.
CC -1- SIMILARITY: BELONGS TO THE STRIATIN FAMILY OF WD-REPEAT PROTEINS.
CC -1- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
DR EMBL: X93326; CAA67701.1; -
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 7.
DR PRINTS: PR00320; GPROTEINBPT.
DR SMART: SM00320; WD40; 5.
DR PROSITE: PS00678; WD_REPEATS_1; 2.
DR PROSITE: PS50082; WD_REPEATS_2; 4.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
DR Calmodulin-binding; Repeat; WD repeat; Coiled coil.
KW DOMAIN 53 120 COILED COIL (POTENTIAL).
FT DOMAIN 149 166 CALMODULIN-BINDING (POTENTIAL).
FT REPEAT 461 500 WD 1.
FT REPEAT 514 553 WD 2.
FT REPEAT 567 606 WD 3.
FT REPEAT 662 701 WD 4.
FT REPEAT 704 743 WD 5.
FT REPEAT 750 779 WD 6.
FT SITE 55 63 CAVEOLIN-BINDING (POTENTIAL).
SQ SEQUENCE 780 AA; 86226 MW; 67F7C2209D560F8 CRC64;

Query Match 7.7%; Score 74; DB 1; Length 780;
Best Local Similarity 23.7%; Pred. No. 17;
Matches 51; Conservative 19; Mismatches 65; Indels 80; Gaps 12;

DB 14 ALNSPLSGMRCIRGADFOCARAVG--LACTFRAFLSSRLQDLYSIVRADRAVPIV 72
DB 398 ALTFPPSSGSKFIMGAD--EALSELDGLGLAG-----LTVANEADSLAYDIA 443
OY 73 NUKDELFPSEWALFSGSGPLKPKARIFSPDGKDLVR--HPTW----- 115
DB 444 NKKDALR--KTNWPKFT-----LRS-----HFDGIRALAPHPLEPVLITASEDHDKMMNL 492
OY 116 -----OKSVWHSQDP-----NGRRL-----TESYCEWRTAPSATG 147
DB 493 OKTAPAKKSTLDVEPIYTFRAHKGFLVLCVWSSNGEQCTSGDTGLIOSWSTTNPNVDP 552
```

```

OY 148 QAS---SILGRLGOS-----ASCHHAYILC 173
DB 553 YDSYDPSVLKPLGHTDVAWGGLAYSANORLLSC 587

RESULT 8
HDAG_HUMAN
ID HDAG_HUMAN STANDARD: PRT; 1215 AA.
AC 09UN27:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Histone deacetylase 6 (HD6).
GN HDAC6 OR JM21
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
(1)
RN SEQUENCE FROM N.A.
RP MEDLINE=99238449; PubMed=10220385;
RA Grozinger C.M., Hassig C.A., Schreiber S.L.;
RT "Three proteins define a class of human histone deacetylases related
to yeast Hda1p."
RL Proc. Natl. Acad. Sci. U.S.A. 96:4868-4873(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strom T.M., Gutwillinger N., Nakamura G., Hellbrand H., Drescher B.,
RA Rosenthal A., Melndl A.;
RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: RESPONSIBLE FOR THE DEACETYLATION OF LYSINE RESIDUES ON
THE N-TERMINAL PART OF THE CORE HISTONES (H2A, H2B, H3 AND H4).
CC HISTONE DEACETYLATION PLAYS AN IMPORTANT ROLE IN TRANSCRIPTIONAL
REGULATION, CELL CYCLE PROGRESSION AND DEVELOPMENTAL EVENTS (BY
SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- SIMILARITY: BELONGS TO THE HISTONE DEACETYLASE / ACUC / APHA
FAMILY. HD SUBFAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF132609; AAD28048.1; -
DR EMBL: AF011972; CAA09893.1; -
DR MIM: 300272; -
DR InterPro: IPR000286; His_deacetylase.
DR InterPro: IPR001607; zf-UBP.
DR Pfam: PF00850; Hist_deacetyl; 2.
DR Pfam: PF02148; zf-UBP; 1.
DR PRINTS: PR01270; HDASUPER.
DR SMART: SM00290; znf-UBP; 1.
KW Hydrolyase; Nuclear protein; Repeat.
FT DOMAIN 87 404 HISTONE DEACETYLASE 1.
FT DOMAIN 482 800 HISTONE DEACETYLASE 2.
SQ SEQUENCE 1215 AA; 131430 MW; 77406633FB11CFA CRC64;

Query Match 7.7%; Score 74; DB 1; Length 1215;
Best Local Similarity 23.3%; Pred. No. 28;
Matches 52; Conservative 18; Mismatches 83; Indels 70; Gaps 10;

OY 5 DDPVPLHLVALN-SPLSGCMRGIRGADPQCGQAR-AVGLAG-----TFPAFL 50
DB 333 EFPQPLVLAAGDALQGDPKGEMAAIPAGFADLTLLMGLAGSKLLSLEGYINERALA 392
OY 51 SSRLODLYSIVRRAD-----RAAVPIVNLKDELLPSPWEALF----- 87

```

```

DB 393 EGVASLHTLL--GDPCMLESPGAPCRSAQASVSCALELLEPWEVLVSTETVERDNM 450
OY 88 -----SGSEGLPKPARIFSPGKDYLRHPWP--QKSWHSGSDNGRRLEISYCTET 138
DB 451 EEDNVESESEEGPWEF-----PVLPLTPVLOSRTGLVYDON---MNNHCNLY 496
OY 139 RTEAPSATGQASSL-----LGRLGOSASCHHAYILC 173
DB 497 DSHHEVFORLIRIMCRLEELGLAGRLCTLTTPRPATEBELTTC 539

RESULT 9
PFX_ECOLI
ID PFX_ECOLI STANDARD: PRT; 512 AA.
AC P29014; P76981;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Exopolysphatase (EC 3.6.1.11) (Exopolypase) (Metaphosphatase).
GN PFX OR B2502 OR Z3765 OR ECS3364.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
RC STRAIN=K12;
RX MEDLINE=93107072; PubMed=8380170;
RA Akiyama M., Crooke E., Kornberg A.;
RT "An exopolysphatase of Escherichia coli. The enzyme and its ppx
gene in a polysphatase operon."
RL J. Biol. Chem. 268:633-639(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / M61655;
RX MEDLINE=97426617; PubMed=92278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Colado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsunashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakanura Y., Nishimoto H.,
RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubdaram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
K-12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features."
RL DNA Res. 4:91-113(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Goebel E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch J.A., Blatner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533(2001).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 050952;

```

RX MEDLINE-21156231; PubMed-11358796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shingawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:111-22(2001).
RN (6)
RP SIMILARITY TO GPPA.
RA MEDLINE-94025037; PubMed-8212131;
RX Reizer J., Reizer A., Sailer M.H. Jr., Bork B., Sander C.;
RT "Exopolyphosphate phosphatase and guanosine pentaphosphate
RT phosphatase belong to the sugar kinase/actin/hsp 70 superfamily.";
RL Trends Biochem. Sci. 18:247-248(1993).
CC -1- FUNCTION: DEGRADATION OF INORGANIC POLYPHOSPHATES. ORTHOPHOSPHATE
CC IS RELEASED PROGRESSIVELY FROM THE ENDS OF POLYPHOSPHATE OF CIRCA
CC 500 RESIDUES LONG. BUT CHAINS OF CIRCA 15 RESIDUES COMPLETE POORLY
CC WITH POLYPHOSPHATE AS SUBSTRATE.
CC -1- CATALYTIC ACTIVITY: [Polyposphate(N) + H(2)O =
CC (polyphosphate)(N-1) + phosphate.
CC -1- COFACTOR: REQUIRES MAGNESIUM FOR ACTIVITY.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Membrane-associated.
CC -1- SIMILARITY: BELONGS TO THE GPPA / PPX FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L06129; AAA24415.1; -
DR EMBL: AE000336; AAC75555.1; -
DR EMBL: D90878; BAA16390.1; -
DR EMBL: D90880; BAA16392.1; -
DR EMBL: AE005479; AAG57612.1; -
DR EMBL: AP002561; BAB36787.1; -
DR PIR: A45333; A45333.
DR Ecogene: EGI1403; ppx.
DR InterPro: IPR003695; ppx_Gppa.
DR Pfam: PF02541; Ppx-Gppa; 1.
KW Hydrolyase; Magnesium; Membrane; Complete proteome.
KW INT_MER 0
FT SEQUENCE 512 AA; 58004 MW; 48611AEF5D9FB9C3 CRC64;

Query Match 7.5%; Score 73; DB 1; Length 512;
Best Local Similarity 21.9%; Pred. No. 13;
Matches 46; Conservative 28; Mismatches 66; Indels 70; Gaps 11;

QY 34 FOGAR-----AVGLAGTFRA-----FLSSRLDLY 58
DB 186 FORARMAAOKLETLTWORRIGQWVNAAGASGITIKAAHEVLMGKEDGITTPEERLEKLV 245
QY 59 -STVRADRAAVPIVNLKDE--LFFPSWEALFSG-----SEGFLPKGARIFSF 103
DB 246 KEVLRHRNFASLSTPLSEBRKTVFVPGGLALIGVFDALAIRELRLSDGALRGV-LYEM 304
QY 104 DGR-----DVLRRPTWPKQSVWMSGDPNGRRL--TESTYCTWTEAPS-ATGQASSLLG- 154
DB 305 EGFRRHODVRSRTASSLANOYHIDSQARVLDITQMYEQWREQPKLAHPQLEALLRW 364
QY 155 -----GRLLGOSASACHAAVIVLCIENS 177
DB 365 AAMLEHVGILNINHSGLHRHSAYI---LONS 391

RESULT 10
YP67_MYCTU STANDARD; PRT; 884 AA.

AC 050654; Q50731.
DT 01-NOV-1997 (Rel. 35, Created)
DR 30-MAY-2000 (Rel. 39, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DT Hypoetical 95.4 kDa protein RV2567.
GN RV2567 OR MT2643 OR MTCY227.34C OR MTCY964.01C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxId=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE-98295967; PubMed-9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. Iii, Tekela F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Emdinova M.D., Salzberg S.L.,
RA Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: SOME, TO M.TUBERCULOSIS RV2411C AND STNECHOCYSTIS PCC
CC 6803 SLD0335.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z77250; CAB01053.1; -
DR EMBL: AE007098; AAK46956.1; -
DR TIGR: MT2643; -
DR Tuberculist: RV2567; -
KW Hypoetical protein; Complete proteome.
KW CONFLICT 645 645 0 -> R (IN REF. 2).
FT SEQUENCE 884 AA; 95448 MW; 95D23A4D2DEB365 CRC64;

Query Match 7.5%; Score 72.5; DB 1; Length 884;
Best Local Similarity 29.5%; Pred. No. 27;
Matches 31; Conservative 15; Mismatches 26; Indels 33; Gaps 6;

QY 63 RADRAVPIVNLKDELLFFPSWEALFSGSEGPLKPKGRIRISFDC-----KDVLRHPW 114
DB 609 RADMIIVA-----PS-TLWSITVDPDPGSLVOSVEELALAAQAVRDOLSDNTW 656
QY 115 -----PKQSVWMSGDPNGRRLTESYCTWTEAPSATGQASSLLG 154
DB 657 MYLANVERAVERHKSDP-----FQSLAE---ADAVLASAQAETLAG 693

RESULT 11
CRK_HUMAN STANDARD; PRT; 304 AA.
AC P46108;

DR PDB: 1CKB; 08-MAY-95.
DR PDB: 1B07; 06-JAN-99.
DR MGI:88508; Crko.
DR InterPro: IPR000980; SH2.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00017; SH3.
DR Pfam: PF00018; SH3.
DR PRINTS: PR00401; SH2DOMAIN.
DR PRINTS: PR00452; SH3DOMAIN.
DR SMART: SM00252; SH2; 1.
DR SMART: SM00326; SH3; 2.
DR PROSITE: PSS0001; SH2; 1.
DR PROSITE: PSS0002; SH3; 1.
DR PROTO-ONCOGENE; SH2 domain; SH3 domain; Repeat; Alternative splicing;
KW Phosphorylation; 3D-structure.
FT DOMAIN 13 118 SH2.
FT DOMAIN 132 192 SH3 1.
FT DOMAIN 256 286 SH3 2.
FT VARSPLIC 205 304 MISSING (IN ISOFORM CRK-I).
SQ SEQUENCE 304 AA; 33814 MW; 5491896FC7A89065 CRC64;

Query Match 7.4%; Score 72; DB 1; Length 304;
Best Local Similarity 27.8%; Pred. No. 8.8;
Matches 20; Conservative 12; Mismatches 24; Indels 16; Gaps 3;

OY 100 IFSFDDK-----DVLRHPTWPKSVHSGSDPNGR--LTESYCEWTREASATG 147
DB 140 LDFPENGDEEDLPFKKGDILIRDKPEEOWMNEDESGKRGMLPVPYEYR----PASA 195

OY 148 QASLLGRLG 159
DB 196 SVSALLGNGEG 207

RESULT 13
CRK_RAT STANDARD; PRT; 304 AA.

AC Q63768;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE PROTO-ONCOGENE C-CRK (P38) (Adapter molecule crk).
GN CRK.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Kidney;
RA MEDLINE-97057214; PubMed-8901553;
RA Kizaka-Kondoh S.; Matsuda M.; Okayama H.;
RT "Crli signals from epidermal growth factor receptor to Ras.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:12177-12182(1996).
CC -1- FUNCTION: THE CRK-I AND CRK-II FORMS DIFFER IN THEIR BIOLOGICAL
CC ACTIVITIES. CRK-II HAS LESS TRANSFORMING ACTIVITY THAN CRK-I. BOTH
CC CRK-I AND CRK-II BIND TO MANY OF TYROSINE-PHOSPHORYLATED PROTEINS
CC THAT BIND TO GRB2.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: CRK-I AND CRK-II (SHOWN HERE);
CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE CRK-I FORM LACKS THE
CC LAST 100 RESIDUES.
CC -1- TISSUE SPECIFICITY: CRK-II IS EXPRESSED IN ALL TISSUES AND CELLS
CC WHEREAS CRK-I IS EXPRESSED AT LOWER LEVEL AND IN LIMITED CELL-
CC TYPES.
CC -1- DOMAIN: THE C-TERMINAL SH3 DOMAIN FUNCTION AS A NEGATIVE MODULATOR
CC FOR TRANSFORMATION AND THE N-TERMINAL SH3 DOMAIN APPEARS TO
CC FUNCTION AS A POSITIVE REGULATOR FOR TRANSFORMATION (BY
CC SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 SH3 DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: D44481; BAA07924.1; -.
DR HSSP: O64010; 1CKA.
DR InterPro: IPR000980; SH2.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00017; SH2; 1.
DR Pfam: PF00018; SH3; 1.
DR PRINTS: PR00401; SH2DOMAIN.
DR PRINTS: PR00452; SH3DOMAIN.
DR SMART: SM00252; SH2; 1.
DR SMART: SM00326; SH3; 2.
DR PROSITE: PSS0001; SH2; 1.
DR PROSITE: PSS0002; SH3; 1.
DR PROTO-ONCOGENE; SH2 domain; SH3 domain; Repeat; Alternative splicing;
KW Phosphorylation.
FT DOMAIN 13 118 SH2.
FT DOMAIN 132 192 SH3 1.
FT DOMAIN 256 286 SH3 2.
FT VARSPLIC 205 304 MISSING (IN ISOFORM CRK-I).
FT VARIANT 244 244 Q -> R (NRK-23 INACTIVE MUTANT).
FT VARIANT 253 253 K -> E (NRK-23 INACTIVE MUTANT).
SQ SEQUENCE 304 AA; 33844 MW; 4CFBF65BE72E265 CRC64;

Query Match 7.4%; Score 72; DB 1; Length 304;
Best Local Similarity 27.8%; Pred. No. 8.8;
Matches 20; Conservative 12; Mismatches 24; Indels 16; Gaps 3;

OY 100 IFSFDDK-----DVLRHPTWPKSVHSGSDPNGR--LTESYCEWTREASATG 147
DB 140 LDFPENGDEEDLPFKKGDILIRDKPEEOWMNEDESGKRGMLPVPYEYR----PASA 195

OY 148 QASLLGRLG 159
DB 196 SVSALLGNGEG 207

RESULT 14

HS75.CANAL STANDARD; PRT; 613 AA.

AC P87232;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Heat shock protein SSB1.
GN SSB1 OR HSP70B.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Microsporid Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WO-1;
RA MEDLINE-97344370; PubMed-9200817;
RA Manen V.; Cervera A.M.; Martinez J.P.; Gozalbo D.;
RT "Molecular cloning of a Candida albicans gene (SSB1) coding for a
RT protein related to the Hsp70 family.";
RL Yeast 13:677-681(1997).
CC -1- FUNCTION: MAY AID IN THE PASSAGE OF THE NASCENT POLYPEPTIDE CHAIN
CC THROUGH THE RIBOSOME CHANNEL INTO THE CYTOSOL. SUCH AN INTERACTION
CC COULD BE CRUCIAL FOR CONTINUOUS TRANSPORT OF THE POLYPEPTIDE;
CC COULD SERVE TO PREVENT THE NASCENT POLYPEPTIDE FROM INTERFERING
CC WITH TRANSLATION BY CLOGGING THE RIBOSOME CHANNEL (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH TRANSLATING
CC RIBOSOMES. MAY BIND DIRECTLY TO THE NASCENT POLYPEPTIDE (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.

```

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X87723; CAA66308.1; -.
CC HSSP: P08107; 1HUO.
CC COMPLEYEAST-2DPAGE; P87222; -.
CC InterPro: IPR001023; HSP70.
CC Pfam: PF00012; HSP70; 1.
CC PRINTS: PR00301; HEATSHOCK70.
CC PROSITE: PS00297; HSP70_1; 1.
CC PROSITE: PS00329; HSP70_2; 1.
CC PROSITE: PS01036; HSP70_3; 1.
CC Heat shock; ATP-binding; Multigene family; Protein biosynthesis.
CC SEQUENCE 613 AA; 66432 MW; 5FBA9F8F99327F9 CRC64;

Query Match 7.4%; Score 72; DB 1; Length 613;
Best Local Similarity 25.6%; Pred. No. 20;
Matches 33; Conservative 20; Mismatches 50; Indels 26; Gaps 6;

QY 44 GFRRAFLSSRLDLYSIVRADRAAVPIVNLKDELFPSEWALFS-----GSGGPL 94
DB 4 GVFQGAIGIDLGTTYSVATYDSAVETIANEQGNRYTPSPVAFTESERLIGDAKKNQAL 63
QY 95 KPGARIF-----SFQKDVLRH-PTWPKSVHSGSDPNGRRLTE-SYCEWTRTEAPS 144
DB 64 NPKNTYFVDAKRLIGRAFDSDSVQKDKISWPKFYV----ESNGQPLIEVEYLDKTKTSPQ 119
QY 145 ATQASSL 153
DB 120 ---EISSMV 125

RESULT 15
NRRL_CHICK STANDARD; PRT; 312 AA.
AC P55806;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, last sequence update)
DT 01-NOV-1997 (Rel. 35, last annotation update)
DE NAD(P)(+)-arginine ADP-ribosyltransferase 1 precursor (EC 2.4.2.31)
DE (mono(ADP-ribosyl)transferase 1) (AT1).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=WHITE LEGHORN; TISSUE=Bone marrow;
RX MEDLINE=95050487; Pubmed=7961658;
RA Tsuchiya M., Hara N., Yamada K., Osago H., Shimoyama M.;
RT "Cloning and expression of cDNA for arginine-specific ADP-
RT ribosyltransferase from chicken bone marrow cells.";
RL J. Biol. Chem. 269:27451-27457(1994).
CC -1- CATALYTIC ACTIVITY: NAD(+) + L-arginine = nicotinamide + N2-
CC (ADP-D-ribosyl)-L-arginine.
CC -1- SUBCELLULAR LOCATION: THE MATURE ENZYME IS PROBABLY SECRETED
CC FROM GRANULOCYTES INTO THE EXTRACELLULAR SPACE.
CC -1- SIMILARITY: BELONGS TO THE ARG-SPECIFIC ADP-RIBOSYLTRANSFERASE
CC FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D31864; BAA06664.1; -.
CC InterPro: IPR000768; ART.
CC Pfam: PF01129; ART; 1.
CC PRINTS: PR00970; RIBTRNSFRASE.
CC PROSITE: PS01291; ART; 1.
CC Transferase; Glycosyltransferase; NAD; Signal; Zymogen.
CC SIGNAL 1 20
CC PROPEP 21 31
CC CHAIN 32 266
CC FT NAD(P)(+)-ARGININE ADP-
CC RIBOSYLTRANSFERASE 1.
CC PROPEP 267 312
CC ACT_SITE 224 224
CC BY SIMILARITY.
CC SEQUENCE 312 AA; 35318 MW; B82360439BC904FC CRC64;

Query Match 7.3%; Score 70.5; DB 1; Length 312;
Best Local Similarity 24.8%; Pred. No. 13;
Matches 30; Conservative 17; Mismatches 45; Indels 29; Gaps 5;

QY 2 SHRDQPVHLVLVNLSPISGGMGIRGADFCQOQARV-----GLAGFRRAFLSS 52
DB 128 SHQDYHSHYHKTLHFLLTQALFALRASQPCRYVYGVGRGIRMTQGRKSVRFQGTST 187
QY 53 RL-----QDLKSYVRADRAAVPIVNL-----KDELFPSEWAL-----FSGSGPL 94
DB 188 SLRREATVNGQDPLLFVYKTC--YGVPIKQSFSPSEDEVLIPPEVEYVINFSDRGSV 245
QY 95 K 95
DB 246 K 246

```

Search completed: August 12, 2002, 10:14:43
 Job time: 130 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 10:11:43 ; Search time 16.76 Seconds
(without alignments)
1049.185 Million cell updates/sec

Title: US-10-080-797-1

Perfect score: 967

Sequence: 1 HSHRDFQPVHLVALNPLS.....SCHHATVLCIENSFWTAKS 183

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	967	100.0	684	2	A53019 collagen alpha 1(X
2	840	86.9	1774	2	B56101 collagen alpha 1(X
3	837	86.6	1315	2	A56101 collagen alpha 1(X
4	529	54.7	1388	2	A53317 collagen alpha 1(X
5	373	38.6	650	2	T22002 hypothetical prote
6	79	8.2	477	2	T05202 pectinesterase hom
7	77	8.0	314	2	F98231 succinoglycan bios
8	77	8.0	314	2	AG3054 succinoglycan bios
9	76	7.9	204	2	B45022 CRK-I - human
10	76	7.9	636	2	JM0047 class I cytochrome
11	75.5	7.8	904	2	T04377 probable alpha-dex
12	75	7.8	351	2	B89781 conserved hypother
13	74.5	7.7	850	2	S56015 gastric mucin MUC5
14	74	7.7	1063	2	T46284 hypothetical prote
15	73.5	7.6	1653	2	B91052 hypothetical prote
16	73.5	7.6	1653	2	G65028 hypothetical prote
17	73.5	7.6	1653	2	F85896 hypothetical prote
18	73	7.5	302	2	C71332 probable smf prote
19	73	7.5	435	2	JC5584 hyaluronoglycosam
20	73	7.5	462	2	T17480 endo-xylanase homo
21	73	7.5	513	1	A45333 exopolyposphatase
22	73	7.5	513	2	D91049 exopolyposphatase
23	73	7.5	513	2	H85893 UDP-N-acetylglucos
24	72.5	7.5	356	2	AH0068 hypothetical prote
25	72.5	7.5	841	2	B71212 hypothetical prote
26	72.5	7.5	884	2	C70729 epidermal growth f
27	72	7.4	239	2	A46243 carbohydrolase kins
28	72	7.4	297	2	AE3128 CRK-II - human
29	72	7.4	304	2	A45022

30	72	7.4	304	2	I58394 c-Crk - mouse
31	72	7.4	319	2	D98159 hypothetical prote
32	72	7.4	494	2	A83204 arginate biosynthe
33	72	7.4	800	2	AG0438 probable Rns acces
34	72	7.4	1006	2	T00050 hypothetical prote
35	71.5	7.4	554	2	T06374 probable pectinest
36	71.5	7.4	554	2	T06468 pectinesterase (EC
37	71.5	7.4	3670	2	T36249 CDA peptide synthet
38	71	7.3	579	2	A70954 hypothetical prote
39	71	7.3	644	2	A72519 probable 2-oxoacid
40	71	7.3	732	2	AF0439 probable Rns acces
41	71	7.3	787	2	G81692 inner membrane pro
42	71	7.3	1154	2	S43275 hypothetical prote
43	70.5	7.3	312	2	A55461 MAD(P)-L-arginine
44	70.5	7.3	628	2	D90677 probable propionyl
45	70.5	7.3	628	2	G85527 probable propionyl

ALIGNMENTS

RESULT 1

A53019 collagen alpha 1(XVII) chain - human (fragment)

N:Contains: endostatin

C:Species: Homo sapiens (man)

C>Date: 18-Aug-1995 #sequence_rev1sion 18-Aug-1995 #text_change 31-Mar-2000

C:Accession: A53019

R:Oh, S.P.; Warman, M.L.; Seldin, M.F.; Cheng, S.D.; Knoll, J.H.M.; Timmons, S.; Olse

Genomics 19, 494-499, 1994

A:Title: Cloning of cDNA and genomic DNA encoding human type XVIII collagen and local

A:Reference number: A53019; MIM:125548

A:Accession: A53019

A:Molecule type: mRNA

A:Residues: 1-684 <OH>

A:Cross-references: GB:125548; NID:9348908; PIRN:AA51864.1; PID:9562794

A:Note: the cited accession number, 125548, is not in Genbank release 103

A:Note: In the authors' translation, 483-Gly is not shown, residues 483-490 are shift

C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit

lated and subsequently O-glycosylated.

C:Comment: Different splice forms of collagen alpha 1(XVII) may be involved in periv

C:Comment: Endostatin is released from collagen alpha 1(XVII) chain by the action of

ay be useful in treating solid tumors.

C:Genetics:

A:Gene: GDB:COL18A1

A:Cross-references: GDB:138752; OMIM:120328

A:Map position: 21q22.3-21q22.3

C:Superfamily: unassigned collagen

C:Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteog

F:1-684/Product: collagen alpha 1(XVII) chain (fragment) #status predicted <MAT>

F:1-59/Domain: collagenous (fragment) #status predicted <CO4>

F:74-115/Domain: collagenous #status predicted <CO5>

F:129-201/Domain: collagenous #status predicted <CO6>

F:212-244/Domain: collagenous #status predicted <CO7>

F:257-278/Domain: collagenous #status predicted <CO8>

F:262-264/Region: cell attachment (R-G-D) motif

F:386-340/Domain: collagenous #status predicted <CO9>

F:354-371/Domain: collagenous #status predicted <CO10>

F:502-684/Product: endostatin #status predicted <EST>

F:509-684/Region: multiplexin collagen carboxyl-terminal similarity

Query Match 100.0%; Score 967; DB 2; Length 684;
Best Local Similarity 100.0%; Pred. No. 3.2e-88;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	HSHRDFQPVHLVALNPLSGMGIRGADPQCFQARAVAGLCTFRAPLSSRLQDIYSI	60
DB	502	HSHRDFQPVHLVALNPLSGMGIRGADPQCFQARAVAGLCTFRAPLSSRLQDIYSI	561
QY	61	VRRADRAAVPIVNLKDELLFPSSWEALFSGSGPLKPARIFSEDGKDVLRHPTWPKSVW	120
DB	562	VRRADRAAVPIVNLKDELLFPSSWEALFSGSGPLKPARIFSEDGKDVLRHPTWPKSVW	621

A:Title: Alpha1(XVII), a collagen chain with frequent interruptions in the collagenous
A:Reference number: A58371; MUID:94240112
A:Accession: A58371
A:Molecule type: mRNA
A:Residues: 1-928 <REH>
A:Cross-references: GB:L16898; NID:9404754; PIDN:AAA37434.1; PID:9553994
R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.
submitted to the EMBL Data Library, August 1993
A:Reference number: S72450
A:Accession: S72450
A:Molecule type: mRNA
A:Residues: 28-687, 'L', 689-734, 'F', 736-751, 'R', 753-1315 <OHM>
A:Cross-references: EMBL:L22545; NID:9348968; PIDN:AA19787.1; PID:9511298
R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994
A:Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa
A:Reference number: A58370; MUID:94240111
A:Accession: S65595
A:Molecule type: mRNA
A:Residues: 28-1315 <OHS>
A:Cross-references: EMBL:L22545
A:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
O-linked and subsequently O-glycosylated.
C:Comment: The different splice forms of collagen alpha 1(XVII) may be involved in per
C:Comment: Endostatin is released from collagen alpha 1(XVII) chain by the action of un
ay be useful in treating solid tumors.
C:Genetics:
A:Gene: MGI:COL18A1
A:Cross-references: MGI:71175
A:Map position: 10:41.0
C:Superfamily: unassigned collagens
C:Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglyc
F:1-25/Domain: signal sequence #status predicted <SIG>
F:24-235/Region: thrombospondin amino-terminal similarity
F:26-1315/Product: collagen alpha 1(XVII) chain, short splice form #status predicted <
F:327-353/Domain: collagenous #status predicted <CO1>
F:364-437/Domain: collagenous #status predicted <CO2>
F:462-583/Domain: collagenous #status predicted <CO3>
F:607-689/Domain: collagenous #status predicted <CO4>
F:704-745/Domain: collagenous #status predicted <CO5>
F:759-831/Domain: collagenous #status predicted <CO6>
F:842-874/Domain: collagenous #status predicted <CO7>
F:887-910/Domain: collagenous #status predicted <CO8>
F:892-894/Region: cell attachment (R-G-D) motif
F:918-969/Domain: collagenous #status predicted <CO9>
F:983-1000/Domain: collagenous #status predicted <CO10>
F:1132-1315/Product: endostatin #status predicted <EST>
F:1133-1315/Region: multiplexin collagen carboxyl-terminal similarity
F:126,488/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:172-228/Disulfide bonds: #status predicted
F:240,245,1257/Binding site: carbohydrate (Ser) (covalent) #status predicted
F:451,454,594/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 86.6%; Score 837; DB 2; Length 1315;
Best Local Similarity 85.1%; Pred. No. 6,4e-75;
Matches 154; Conservative 16; Mismatches 11; Indels 0; Gaps 0;

DB 1 HSHRDQPVLAHVALNSPLSGRGIRGADFOCFQARAVGLAGTPRAFLSSRLQDLYST 60
HSHRDQPVLAHVALNSPLSGRGIRGADFOCFQARAVGLAGTPRAFLSSRLQDLYST 1191
DB 1132 HTIQDQPVLAHVALNSPLSGRGIRGADFOCFQARAVGLAGTPRAFLSSRLQDLYST 1191
HTIQDQPVLAHVALNSPLSGRGIRGADFOCFQARAVGLAGTPRAFLSSRLQDLYST 1191

DB 61 VRRADRAVPVIVLKDELFPSSWEALFSGSEGLPKGARIFSDGKVDLHPTWPKSVW 120
VRRADRAVPVIVLKDELFPSSWEALFSGSEGLPKGARIFSDGKVDLHPTWPKSVW 120
DB 1192 VRRADRAVPVIVLKDELFPSSWEALFSGSEGLPKGARIFSDGKVDLHPTWPKSVW 1251
VRRADRAVPVIVLKDELFPSSWEALFSGSEGLPKGARIFSDGKVDLHPTWPKSVW 1251

DB 121 HGSDFNGRRLTESYCTWRTAPATGQASLLGRLGSAASHAHYVILCIENSFM 180
HGSDFNGRRLTESYCTWRTAPATGQASLLGRLGSAASHAHYVILCIENSFM 180
DB 1252 HGSDFNGRRLTESYCTWRTAPATGQASLLGRLGSAASHAHYVILCIENSFM 1311
HGSDFNGRRLTESYCTWRTAPATGQASLLGRLGSAASHAHYVILCIENSFM 1311

DB 181 A 181
181 A 181
DB 1312 S 1312
1312 S 1312

RESULT 4
A53317
collagen alpha 1(XV) chain precursor - human
N:Alternate names: procollagen alpha 1(XV) chain
C:Species: Homo sapiens (hmn)
C>Date: 07-Jul-1995 #sequence
C:Accession: A53317; S28778
R:Kivirikko, S.; Heinaemeaki, P.; Rehn, M.; Honkanen, N.; Myers, J.C.; Pihlajaniemi, J. Biol. Chem. 269, 4773-4779, 1994
A:Title: Primary structure of the alpha1 chain of human type XV collagen and exon-int
A:Reference number: A53317; MUID:94148920
A:Accession: A53317
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1388 <KIT>
A:Cross-references: GB:L25280
A:Note: nucleotide sequence and conceptual translation not complete
R:Muragaki, Y.; Abe, N.; Minomiy, Y.; Olsen, B.R.; Ooshima, A.
J. Biol. Chem. 269, 4042-4046, 1994
A:Title: The human alpha1(XV) collagen chain contains a large amino-terminal non-trip
A:Reference number: A53146; MUID:94140817
A:Accession: A53146
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-9, 'S', 'I', '48', 'V', '50-94', 'A', '96-149', 'A', '151-203', 'V', '205-408', 'A', '410-569' <M
A:Cross-references: GB:D2130; NID:9415605; PIDN:BA04762.1; PID:d1005294; PID:946070
R:Myers, J.C.; Kivirikko, S.; Gordon, M.K.; Dion, A.S.; Pihlajaniemi, T.
Proc. Natl. Acad. Sci. U.S.A. 89, 10144-10148, 1992
A:Title: Identification of a previously unknown human collagen chain, alpha1(XV), cha
A:Reference number: S28778; MUID:93066196
A:Accession: S28778
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 544-640, 'P', 642-811, 'P', 813-1252 <MYE>
C:Genetics:
A:Gene: GDB:COL15A1
A:Cross-references: GDB:132578; OMIM:120325
A:Map position: 9q21-9q22
C:Superfamily: unassigned collagens
C:Keywords: signal sequence #status predicted <SIG>
F:1-22/Domain: signal sequence #status predicted <MAT>
F:23-1388/Product: collagen alpha 1(XV) chain #status predicted <MAT>
F:1216-1388/Region: multiplexin collagen carboxyl-terminal similarity

Query Match 54.7%; Score 529; DB 2; Length 1388;
Best Local Similarity 55.9%; Pred. No. 3,4e-44;
Matches 99; Conservative 30; Mismatches 44; Indels 4; Gaps 1;

DB 7 QPVHVALNSPLSGRGIRGADFOCFQARAVGLAGTPRAFLSSRLQDLYSTVRRADR 66
QPVHVALNSPLSGRGIRGADFOCFQARAVGLAGTPRAFLSSRLQDLYSTVRRADR 1270
DB 1215 KPAHLAALNPFSGDIR---ADFOCFQARAVGLAGTPRAFLSSRLQDLYSTVRRADR 1270
KPAHLAALNPFSGDIR---ADFOCFQARAVGLAGTPRAFLSSRLQDLYSTVRRADR 1270

DB 67 AAVPVLKDELFPSSWEALFSGSEGLPKGARIFSDGKVDLHPTWPKSVW 126
AAVPVLKDELFPSSWEALFSGSEGLPKGARIFSDGKVDLHPTWPKSVW 126
DB 1271 YSLPVLKQGVLENMNDISFSGHGGFNNHPIRIFSDGRIMDPEMPQVIMHSGSPH 1330
YSLPVLKQGVLENMNDISFSGHGGFNNHPIRIFSDGRIMDPEMPQVIMHSGSPH 1330

DB 127 GRRILTESYCTWRTAPATGQASLLGRLGSAASHAHYVILCIENSFM 183
GRRILTESYCTWRTAPATGQASLLGRLGSAASHAHYVILCIENSFM 183
DB 1331 GVRILVDNYCEAWRYADRAVAVGLASPLSTGKILDKAKVSCANRLIVLCIENSFM 1387
GVRILVDNYCEAWRYADRAVAVGLASPLSTGKILDKAKVSCANRLIVLCIENSFM 1387

RESULT 5
T22002
hypothetical protein F39H11.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence
C:Accession: T22002
R:White, S.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19500
A:Accession: T22002

R:Matsumura, M.; Tanaka, S.; Nagata, S.; Kojima, A.; Kurata, T.; Shibuya, M.
Mol. Cell. Biol. 12, 3482-3489, 1992
A:Title: Two species of human CRK cDNA encode proteins with distinct biological activities
A:Reference number: A45022; MIM:92334347
A:Accession: B45022
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-204 <MAT>
A:Experimental source: embryonic lung cells
A:Note: sequence extracted from NCBI backbone (NCBI:108771, NCBI:108772)
C:Superfamily: crk transforming protein; SH2 homology; SH3 homology
F:13-118/Domain: SH2 homology <SH2>
F:139-187/Domain: SH3 homology <SH3>

Query Match 7.9%; Score 76; DB 2; Length 204;
Best Local Similarity 29.0%; Pred. No. 4.6;
Matches 20; Conservative 12; Mismatches 21; Indels 16; Gaps 3;

QY 100 IFSPDK-----DVLHPTWPOKSWHSGDPNGRR--LTESYCEWTREAPSGATG 147
DB 140 LFPNGNDEDELPFKKGDILIRIDKPEQWMAEDSGKGMIPVYEKR-----PASA 195
QY 148 QASLLGR 156
DB 196 SVSALIGR 204

RESULT 10
JM0047
Class I cytokine receptor precursor - human
N:Alternate names: MSX-1
C:Species: Homo sapiens (man)
C:Date: 17-Jun-1998 #sequence_revision 10-Jul-1998 #text_change 21-Jul-2000
C:Accession: JM0047
R:Spieker, C.A.; Grant, F.J.; Baumgartner, J.W.; Presnell, S.R.; Schrader, S.K.; Yamaguchi, R.; Blochem, B.; Res. Commun. 246, 82-90, 1998
A:Title: Cloning and characterization of a novel class I cytokine receptor.
A:Reference number: JM0047; MIM:98262921
A:Accession: JM0047
A:Molecule type: mRNA
A:Residues: 1-636 <SPR>
A:Cross-references: GB:AF053004; NID:g3153240; PIDN:AAC39755.1; PID:g3153241
A:Experimental source: brain
C:Genetics:
A:Map position: 19p13.11
C:Keywords: glycoprotein
F:1-33/Domain: signal sequence #status predicted <SIG>
F:515-540/Domain: transmembrane #status predicted <TM>
F:554-561/Domain: cytoplasmic #status predicted <CTP>
F:51,76,302,311,374,382,467/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 7.9%; Score 76; DB 2; Length 636;
Best Local Similarity 24.2%; Pred. No. 18;
Matches 39; Conservative 12; Mismatches 46; Indels 64; Gaps 7;

QY 23 MRIRADP-----OCFOARAVGLAGTFRAFLSSRLDL----- 57
DB 1 MRGRGAPFWLWLPKIALPLLVLFQRTPGSAGAPLOCGYGVGLDNCMEPLGLD 60
QY 58 -----YSIVRRADRAAVPI-----VNLKDELLF-----PSWEALFS 88
DB 61 GAPSEHLLOSQRKSKTKQVAVAAAGRSVAITREQLTHSDKLLWGTAKAGPLMPVEFV 120
QY 89 GSEGLPKGA-RI--FSPDKDVLK-----HPTWPOKSV 119
DB 121 NLETKMKNAPNAPRLGPDVDFSEDDPLEATVHMAPPTWPSHKV 161

RESULT 11
T04377
probable alpha-dextrin endo-1,6-alpha-glucosidase (EC 3.2.1.41) - barley

N:Alternate names: pullulanase
C:Species: Hordeum vulgare (barley)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 08-Oct-1999
C:Accession: T04377
R:Lox, F.; Kristensen, M.; Planchot, V.; Leah, R.; Svendsen, I.; Svenson, B.
submitted to the EMBL Data Library, December 1997
A:Description: Isolation and characterization of starch debranching enzyme, limit dex
A:Reference number: Z15320
A:Accession: T04377
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-904 <LOK>
A:Cross-references: EMBL:AF022725; NID:g2502057; PIDN:AD04189.1; PID:g2677837
A:Experimental source: cv. Igr1
C:Genetics:
A:Gene: HVD99
A:Introns: 21/3; 70/1; 87/3; 124/1; 169/3; 235/2; 285/2; 305/3; 346/1; 376/2; 407/3;
C:Keywords: glycosidase; hydrolase

Query Match 7.8%; Score 75.5; DB 2; Length 904;
Best Local Similarity 22.5%; Pred. No. 31;
Matches 45; Conservative 19; Mismatches 67; Indels 69; Gaps 8;

QY 17 SPLSGMRGIRGADPOCFQOARAVGLAGT-----PRAFLSSRLDYSIVR---- 62
DB 62 SPNSG---GIQYDSKELDPESAGLEPTVYQKRPFTSSYAFVPSVAVASLVKQOLV 118
QY 63 ----RADRAVPIVNL-----KDELLFSPWE-----ALFSG 89
DB 119 VASFGADGKHVDVYGLQPLGVLDMPFAYTGPVAGVSEDSVSLHLMPTAGVSVCFPDG 178
QY 90 SEGP-----LKPARIFFSPDK-----DVLHPTWPOKSWHSGDPNGRRLT 131
DB 179 PAGALFTVYDLKESNGVSVTGPPEMENRYLYLEVIVY-HPTKAQVLCIAGDPYARSL 237
QY 132 ESYCEWTREAPSAVGOASS 151
DB 238 ANGARTLVLDINNETLKPAS 257

RESULT 12
B89781
Conserved hypothetical protein SA0184 [Imported] - staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: B89781
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O
ma, A.; Mizutani-U, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratsatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MIM:21311952; PMID:11418146
A:Accession: B89781
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1351 <KUR>
A:Cross-references: GB:BA000018; PID:g13700106; PIDN:BA041405.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA0184

Query Match 7.8%; Score 75; DB 2; Length 351;
Best Local Similarity 23.3%; Pred. No. 11;
Matches 42; Conservative 17; Mismatches 67; Indels 54; Gaps 8;

QY 23 MRIRGADPOCFQOARAVGLAGTFRAFLSSRLDL-----YSIVRRADRA-AVPIV 72
DB 97 IEAIMAOGKCCCLNASTIS-----RELLTSLDQDLNFTLLSCFNHYVPRDGLSLV 151
QY 73 NLKDELLFSPWEALFSGSEGLPKGARIFSPDKDVLHPTWPOKSWHSGDPNGRRLTE 132

Db 152 NKNNELLY-----QFNPKAQIYGFIVSGSLRGRL-----HKGLPT----- 186

Qy 133 SYCETWRTAP-----SATGOASSLGGRLGOSAS-----CHNAVILCIENSFMT 180

Db 187 --TEATRHSHPVAAKLLQETGVSEVLVGDLSILEMROAKQIDLCFCHRHFTLCIEEVFDI 244

RESULT 13

556015

gastric mucin MUC5AC - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 20-Apr-2000

C:Accession: S56015; S53361

R:Klomp, L.W.J.; van Kers, L.; Strous, G.J.

Biochem. J. 308, 831-838, 1995

A:Title: Cloning and analysis of human gastric mucin cDNA reveals two types of conserved

A:Reference number: S56015; MUID:97104281

A:Accession: S56015

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-850 <KLO>

A:Cross-references: EMBL:X81649; NID:9547516; PIDN:CA57309.1; PID:9547517

R:Guyonnet-Duperrat, V.; Audie, J.P.; Debailleul, V.; Laine, A.; Bulsine, M.P.; Gallegue-Biochem. J. 305, 211-219, 1995

A:Title: Characterization of the human mucin gene MUC5AC: a consensus cysteine-rich dome

A:Reference number: S53361; MUID:95126907

A:Accession: S53361

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 648-678, 'L', 680-733, 'L', 735-760 <GUY>

A:Cross-references: EMBL:Z34280; NID:9563380; PIDN:CA84034.1; PID:9563381

A:Experimental source: clone J0J32

A:Note: this publication is not cited in GenBank entry HSMUCIN5, release 113.0

Query Match 7.7% Score 74.5; DB 2; Length 850;

Best Local Similarity 24.7% Pred. No. 37;

Matches 39; Conservative 23; Mismatches 71; Indels 25; Gaps 6;

Qy 25 GIRGADFQCFQOARAVGLAGTFRAPFLSSRLQDLSIVRRAD-AAVPIVNLKDELFPSPW 83

Db 592 GINGCDFDFQKLRDEGY--TF-----CESPRSYQCARAESPPNPLADLGQDVICSHT 642

Qy 84 EALFSGSEGPLK-----GARIFSEPDGKDVLRHPTWPKSVW-----HGSDDNGRRRLTE 132

Db 643 EELICLNKQRLPICYNYEIRIQCCETVWCRDITRPKTVATTRPTGAGQTQTFET 702

Qy 133 SYCETWRTAPSATGO-----ASSLLGGRLLGOSASACH 166

Db 703 TIMPSASTEQPTATSRGCPATSVTQGTHTTPTVTRNCH 740

RESULT 14

746284

hypothetical protein DKFZp566E044.1 - human

C:Species: Homo sapiens (man)

C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000

C:Accession: T46284

R:Blöcker, H.; Boecher, M.; Brandt, P.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

A:Title: submitted to the Protein Sequence Database, January 2000

A:Reference number: Z23032

A:Accession: T46284

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1063 <AAA>

A:Cross-references: EMBL:AL137696

A:Experimental source: fetal kidney; clone DKFZp566E044

C:Genetics:

A:Note: DKFZp566E044.1

Query Match 7.7% Score 74; DB 2; Length 1063;

Best Local Similarity 23.3% Pred. No. 54;

Matches 52; Conservative 18; Mismatches 83; Indels 70; Gaps 10;

Qy 5 DFGQVLHVALN--SPRSGMKGIRGADFQCFQAR--AVGLAG-----TFRAPL 50

Db 181 EFGQVLVLAAGDALQGGPKGEMATPAGLQTLNLGLAGSKRLISLEGYNRLALA 240

Qy 51 SSRLDLYSLVRAD-----RAAVPIVNLKDELFPSPWALF----- 87

Db 241 EGVASALHTLL--GDCPMLLESPGAPCRSAQTSVSCALELPEFWEVLVRSSTETVERDM 298

Qy 88 -----SGSEGPLKPGARLFSPDGKDVLRHPTW--QKSVHMGSDPNGRRLTESYCEW 138

Db 299 EEDNVVESEEGPWEF-----PVLPILTVPVLOSRITGLVTDQ-----MANHCNLM 344

Qy 139 RREPASATGOASSL-----LGRLLGOSASACHNAVILVC 173

Db 345 DSHNRPVQRIILRMCRLEBELGLAGRLCLTLTPRPATEALLTC 387

RESULT 15

B91052

hypothetical protein ECS3386 [imported] - Escherichia coli (strain O157:H7, substrain B91052)

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C:Accession: B91052

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: B91052

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1653 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA836809.1; PID:913362856; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RMD 0509952

C:Genetics:

A:Gene: ECS3386

Query Match 7.6% Score 73.5; DB 2; Length 1653;

Best Local Similarity 24.4% Pred. No. 1e+02;

Matches 32; Conservative 21; Mismatches 47; Indels 31; Gaps 5;

Qy 41 GLAGTFRAPFLSSRLQDLY-----SIVRRADRAAVPIVNLKDELFPSPWALFSGSE 91

Db 380 GARGYSKPFMFGRDLYRPGETVILNGLRDADKALNPRIKLDVIRPDQVLRSVYS 439

Qy 92 GPLKPGARLFSPDGKDVLRHPTWPKS-----VWH-----GSDRNGRRRLTESYCEWRT-- 141

Db 440 QP-----ENGLYHFTWPLDLSNATGMWHLIRANTGDNQYRMDFHVEDFWEFERM 487

Qy 142 APSATGOASSL 152

Db 488 ALNLIGSEKTPFL 498

Search completed: August 12, 2002, 10:13:51

Job time: 128 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 10:11:08 ; Search time 13.23 seconds
(Without alignments)
337.860 Million cell updates/sec

Title: US-10-080-797-1
Perfect score: 967
Sequence: 1 HSHRDFQVFLHLVNLSP...SCHNAVYIVLCIENSPMTASK 183

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued Patents-AA:*
2: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6A_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/6B_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/6C_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	967	100.0	183	US-09-206-059-2	Sequence 2, Appl
2	962	99.5	182	US-09-561-500-14	Sequence 14, Appl
3	962	99.5	182	US-09-561-108-14	Sequence 14, Appl
4	962	99.5	182	US-09-315-689-3	Sequence 3, Appl
5	937	96.9	178	US-09-315-689-5	Sequence 5, Appl
6	840	86.9	191	US-09-561-500-13	Sequence 13, Appl
7	840	86.9	191	US-09-561-108-13	Sequence 13, Appl
8	837	86.6	195	US-08-159-784-2	Sequence 2, Appl
9	796	82.3	185	US-08-985-526-36	Sequence 36, Appl
10	516	53.4	191	US-08-159-784-3	Sequence 3, Appl
11	160	16.5	35	US-09-046-985-2	Sequence 2, Appl
12	160	16.5	35	US-09-474-743-2	Sequence 2, Appl
13	101	10.4	22	US-09-046-985-7	Sequence 7, Appl
14	101	10.4	22	US-09-474-743-7	Sequence 7, Appl
15	101	10.4	22	US-08-740-168A-1	Sequence 1, Appl
16	96	9.9	20	US-09-349-429-1	Sequence 1, Appl
17	96	9.9	20	US-09-315-689-1	Sequence 1, Appl
18	94	9.7	16	US-09-385-442-32	Sequence 32, Appl
19	76	7.9	16	US-08-653-740-3	Sequence 3, Appl
20	76	7.9	16	US-09-073-594-3	Sequence 3, Appl
21	76	7.9	16	US-09-275-925-3	Sequence 3, Appl
22	76	7.9	16	US-08-653-740-5	Sequence 5, Appl
23	76	7.9	16	US-09-073-594-5	Sequence 5, Appl
24	76	7.9	16	US-09-275-925-5	Sequence 5, Appl
25	74	7.7	256	US-07-906-349A-8	Sequence 8, Appl
26	74	7.7	256	US-08-167-035-4	Sequence 4, Appl
27	74	7.7	256	US-08-208-887A-4	Sequence 4, Appl

28	74	7.7	256	US-08-539-005-4	Sequence 4, Appl
29	73	7.5	311	US-08-987-743-2	Sequence 2, Appl
30	73	7.5	435	US-08-733-360A-1	Sequence 1, Appl
31	73	7.5	435	US-08-733-360A-3	Sequence 3, Appl
32	73	7.5	435	US-08-987-743-6	Sequence 6, Appl
33	73	7.5	435	US-08-987-743-15	Sequence 15, Appl
34	73	7.5	435	US-08-916-935-1	Sequence 1, Appl
35	73	7.5	435	US-08-916-935-3	Sequence 3, Appl
36	72	7.4	304	US-08-630-915A-28	Sequence 28, Appl
37	70	7.2	210	US-09-188-930-148	Sequence 148, Appl
38	70	7.2	6095	US-09-144-085-2	Sequence 2, Appl
39	69	7.1	190	US-09-046-985-15	Sequence 15, Appl
40	69	7.1	190	US-09-474-743-15	Sequence 15, Appl
41	69	7.1	313	US-09-347-803-6	Sequence 6, Appl
42	69	7.1	327	5171684-6	Patent No. 5171684
43	68.5	7.1	1289	US-07-876-280-4	Sequence 4, Appl
44	68.5	7.1	1289	US-07-675-772-4	Sequence 4, Appl
45	68.5	7.1	1289	US-08-063-170-4	Sequence 4, Appl

ALIGNMENTS

```
RESULT 1
US-09-206-059-2
: Sequence 2, Application US/09206059
: Patent No. 6201104
: GENERAL INFORMATION:
: APPLICANT: Macdonald, Nicholas
: APPLICANT: Sim, Kim Lee
: TITLE OF INVENTION: Angiogenesis-Inhibiting Protein Binding Peptides and
: FILE REFERENCE: 05213-0370
: CURRENT APPLICATION NUMBER: US/09/206,059
: NUMBER OF SEQ ID NOS: 80
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 2
: LENGTH: 183
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-206-059-2

Query Match      100.0%; Score 967; DB 4; Length 183;
Best Local Similarity 100.0%; Pred. No. 2e+110;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 HSHRDFQVFLHLVNLSPGSGMGIRGADFOCQOARAVGLAGTFRFLSSRLQDLYSI 60
Db      1 HSHRDFQVFLHLVNLSPGSGMGIRGADFOCQOARAVGLAGTFRFLSSRLQDLYSI 60

QY      61 VRRADRAVPIVNLKDELLFSPWEALFSGSEGPLKPGARIFSPDKDYLRRPTWPQKSVW 120
Db      61 VRRADRAVPIVNLKDELLFSPWEALFSGSEGPLKPGARIFSPDKDYLRRPTWPQKSVW 120

QY      121 HGSDPNGRRLTESYCEWTREAPSATGQASSLLGGRLLGQSAASHNAVYIVLCIENSPMT 180
Db      121 HGSDPNGRRLTESYCEWTREAPSATGQASSLLGGRLLGQSAASHNAVYIVLCIENSPMT 180

QY      181 ASK 183
Db      181 ASK 183

RESULT 2
US-09-561-500-14
: Sequence 14, Application US/09561500
: Patent No. 6342219
: GENERAL INFORMATION:
: APPLICANT: Philip E. Thorpe
: APPLICANT: Rolf A. Brecken
: TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
```


FILE REFERENCE: 4001.002500
CURRENT APPLICATION NUMBER: US/09/561,500
CURRENT FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/131,432
PRIOR FILING DATE: 1999-04-28
NUMBER OF SEQ ID NOS: 44
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 14
LENGTH: 182
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-500-14

Query Match 99.5%; Score 962; DB 4; Length 182;
Best Local Similarity 100.0%; Pred. No. 8.2e-110;
Matches 182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HSHRDQPVHLVALNSPLSGMKGIRGADFCFOQARAVGLAGTFRATLSSRLDLYSI 60
DB 1 HSHRDQPVHLVALNSPLSGMKGIRGADFCFOQARAVGLAGTFRATLSSRLDLYSI 60
QY 61 VRRADRAAVPIVNLKDELLFPSSWEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVW 120
DB 61 VRRADRAAVPIVNLKDELLFPSSWEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVW 120
QY 121 HGSDPNGRRLTESYCEWTREAPSATGQASSLLGRLGQSAASHHAYIVLCIENSFMT 180
DB 121 HGSDPNGRRLTESYCEWTREAPSATGQASSLLGRLGQSAASHHAYIVLCIENSFMT 180
QY 181 AS 182
DB 181 AS 182

RESULT 3
US-09-561-108-14
Sequence 14, Application US/09561108
Patent No. 6342221
GENERAL INFORMATION:
APPLICANT: Phillip E. Thorpe
APPLICANT: Rolf A. Brecken
TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
FILE REFERENCE: 4001.002584
CURRENT APPLICATION NUMBER: US/09/561,108
CURRENT FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/131,432
PRIOR FILING DATE: 1999-04-28
NUMBER OF SEQ ID NOS: 44
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 14
LENGTH: 182
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-108-14

Query Match 99.5%; Score 962; DB 4; Length 182;
Best Local Similarity 100.0%; Pred. No. 8.2e-110;
Matches 182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HSHRDQPVHLVALNSPLSGMKGIRGADFCFOQARAVGLAGTFRATLSSRLDLYSI 60
DB 1 HSHRDQPVHLVALNSPLSGMKGIRGADFCFOQARAVGLAGTFRATLSSRLDLYSI 60
QY 61 VRRADRAAVPIVNLKDELLFPSSWEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVW 120
DB 61 VRRADRAAVPIVNLKDELLFPSSWEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVW 120

DB 61 VRRADRAAVPIVNLKDELLFPSSWEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVW 120
QY 121 HGSDPNGRRLTESYCEWTREAPSATGQASSLLGRLGQSAASHHAYIVLCIENSFMT 180
DB 121 HGSDPNGRRLTESYCEWTREAPSATGQASSLLGRLGQSAASHHAYIVLCIENSFMT 180
QY 181 AS 182
DB 181 AS 182

RESULT 4
US-09-315-689-3
Sequence 3, Application US/09315689
Patent No. 6346510
GENERAL INFORMATION:
APPLICANT: Folkman, Judah
APPLICANT: O'Reilly, Michael
TITLE OF INVENTION: Therapeutic Antiangiogenic Endostatin Compositions
FILE REFERENCE: 05213-0229
CURRENT APPLICATION NUMBER: US/09/315,689
CURRENT FILING DATE: 1999-05-20
NUMBER OF SEQ ID NOS: 6
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 3
LENGTH: 182
TYPE: PRT
ORGANISM: Homo sapiens
US-09-315-689-3

Query Match 99.5%; Score 962; DB 4; Length 182;
Best Local Similarity 100.0%; Pred. No. 8.2e-110;
Matches 182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HSHRDQPVHLVALNSPLSGMKGIRGADFCFOQARAVGLAGTFRATLSSRLDLYSI 60
DB 1 HSHRDQPVHLVALNSPLSGMKGIRGADFCFOQARAVGLAGTFRATLSSRLDLYSI 60
QY 61 VRRADRAAVPIVNLKDELLFPSSWEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVW 120
DB 61 VRRADRAAVPIVNLKDELLFPSSWEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVW 120
QY 121 HGSDPNGRRLTESYCEWTREAPSATGQASSLLGRLGQSAASHHAYIVLCIENSFMT 180
DB 121 HGSDPNGRRLTESYCEWTREAPSATGQASSLLGRLGQSAASHHAYIVLCIENSFMT 180
QY 181 AS 182
DB 181 AS 182

RESULT 5
US-09-315-689-5
Sequence 5, Application US/09315689
Patent No. 6346510
GENERAL INFORMATION:
APPLICANT: Folkman, Judah
APPLICANT: O'Reilly, Michael
TITLE OF INVENTION: Therapeutic Antiangiogenic Endostatin Compositions
FILE REFERENCE: 05213-0229
CURRENT APPLICATION NUMBER: US/09/315,689
CURRENT FILING DATE: 1999-05-20
NUMBER OF SEQ ID NOS: 6
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 5
LENGTH: 178
TYPE: PRT
ORGANISM: Homo sapiens
US-09-315-689-5

Query Match 96.9%; Score 937; DB 4; Length 178;

Best Local Similarity 100.0%; Pred. No. 9.1e-107;
Matches 178; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DFOVPLHLVALNSPLSGMGRIGADFCFOQARAAGLAGTFRFAFLSSRLQDLYSI 64
Db 1 DFOVPLHLVALNSPLSGMGRIGADFCFOQARAAGLAGTFRFAFLSSRLQDLYSI 60
QY 65 DRAVPIVNLKDELFPSPWALFSGSEGPLKPGARIFSPDGKDYLRHPTWPKSVHMSD 124
Db 61 DRAVPIVNLKDELFPSPWALFSGSEGPLKPGARIFSPDGKDYLRHPTWPKSVHMSD 120
QY 125 PNGRLTESYCEWTREAPATGQASSLLGRLIGQSAASCHHAYIVLCIENSFMTAS 182
Db 121 PNGRLTESYCEWTREAPATGQASSLLGRLIGQSAASCHHAYIVLCIENSFMTAS 178

RESULT 6
US-09-561-500-13
Sequence 13, Application US/09561500
Patent No. 6342219

GENERAL INFORMATION:
APPLICANT: Philip E. Thorpe
TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
FILE REFERENCE: 4001.002500
CURRENT APPLICATION NUMBER: US/09/561,500
CURRENT FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/131,432
NUMBER OF SEQ ID NOS: 44
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 13
LENGTH: 191
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-500-13

Query Match 86.9%; Score 840; DB 4; Length 191;
Best Local Similarity 85.6%; Pred. No. 7.5e-95;

Matches 155; Conservative 15; Mismatches 11; Indels 0; Gaps 0;

QY 1 HSHRDFOPVHLVALNSPLSGMGRIGADFCFOQARAAGLAGTFRFAFLSSRLQDLYSI 60
Db 8 HSHRDFOPVHLVALNSPLSGMGRIGADFCFOQARAAGLAGTFRFAFLSSRLQDLYSI 67
QY 61 VRRADRAAVPIVNLKDELFPSPWALFSGSEGPLKPGARIFSPDGKDYLRHPTWPKSV 120
Db 68 VRRADRAAVPIVNLKDELFPSPWALFSGSEGPLKPGARIFSPDGKDYLRHPTWPKSV 127
QY 121 HGSDFPGRRLTESYCEWTREAPATGQASSLLGRLIGQSAASCHHAYIVLCIENSFMT 180
Db 128 HGSDFPGRRLTESYCEWTREAPATGQASSLLGRLIGQSAASCHHAYIVLCIENSFMT 187
QY 181 A 181
Db 188 S 188

RESULT 7

US-09-561-108-13
Sequence 13, Application US/09561108
Patent No. 6342221

GENERAL INFORMATION:
APPLICANT: Philip E. Thorpe
TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
FILE REFERENCE: 4001.002584
CURRENT APPLICATION NUMBER: US/09/561,108
CURRENT FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/131,432

PRIOR FILING DATE: 1999-04-28
NUMBER OF SEQ ID NOS: 44
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 13
LENGTH: 191
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-108-13

Query Match 86.9%; Score 840; DB 4; Length 191;
Best Local Similarity 85.6%; Pred. No. 7.5e-95;
Matches 155; Conservative 15; Mismatches 11; Indels 0; Gaps 0;

QY 1 HSHRDFOPVHLVALNSPLSGMGRIGADFCFOQARAAGLAGTFRFAFLSSRLQDLYSI 60
Db 8 HSHRDFOPVHLVALNSPLSGMGRIGADFCFOQARAAGLAGTFRFAFLSSRLQDLYSI 67
QY 61 VRRADRAAVPIVNLKDELFPSPWALFSGSEGPLKPGARIFSPDGKDYLRHPTWPKSV 120
Db 68 VRRADRAAVPIVNLKDELFPSPWALFSGSEGPLKPGARIFSPDGKDYLRHPTWPKSV 127
QY 121 HGSDFPGRRLTESYCEWTREAPATGQASSLLGRLIGQSAASCHHAYIVLCIENSFMT 180
Db 128 HGSDFPGRRLTESYCEWTREAPATGQASSLLGRLIGQSAASCHHAYIVLCIENSFMT 187
QY 181 A 181
Db 188 S 188

RESULT 8
US-08-159-784-2
Sequence 2, Application US/08159784
Patent No. 5643783

GENERAL INFORMATION:
APPLICANT: Bjorn R. Olsen
TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/159,784
FILING DATE: December 1, 1993
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: John F. Freeman
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 00246/170001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 195
TYPE: amino acid
STRANDEDNESS: N/A

Db 134 GVLVNDYCEAMRTADTAVTGLASPLSTGKILDKAKYSCANRLIVICIENSFMTDAR 190

RESULT 11

US-09-046-985-2

Sequence 2, Application US/09046985

Patent No. 6121236

GENERAL INFORMATION:

APPLICANT: Ben-Sasson, Shmuel A.

TITLE OF INVENTION: NOVEL MULTIVALENT LIGANDS WHICH MODULATE

TITLE OF INVENTION: ANGIOGENESIS

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: Two Millitia Drive

CITY: Lexington

STATE: Massachusetts

COUNTRY: USA

ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/046,985

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Brook, David E.

REGISTRATION NUMBER: 22,592

REFERENCE/DOCKET NUMBER: CMCC-614

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781) 861-9540

TELEFAX: (781) 861-6240

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 35 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-09-046-985-2

Query Match 16.5%; Score 160; DB 3; Length 35;

Best Local Similarity 100.0%; Pred. No. 9.9e-13;

Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 36 QARAVGLAGTFRAPFLSRLODLYSIVRRADRAAV 69

Db 1 QARAVGLAGTFRAPFLSRLODLYSIVRRADRAAV 34

RESULT 12

US-09-474-743-2

Sequence 2, Application US/09474743

Patent No. 6255716

GENERAL INFORMATION:

APPLICANT: Ben-Sasson, Shmuel A.

TITLE OF INVENTION: NOVEL MULTIVALENT LIGANDS WHICH MODULATE

TITLE OF INVENTION: ANGIOGENESIS

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: Two Millitia Drive

CITY: Lexington

STATE: Massachusetts

COUNTRY: USA

ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/474,743

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION NUMBER:

APPLICATION NUMBER: 09/046,985

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Brook, David E.

REGISTRATION NUMBER: 22,592

REFERENCE/DOCKET NUMBER: CMCC-614

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781) 861-9540

TELEFAX: (781) 861-6240

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 35 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-09-474-743-2

Query Match 16.5%; Score 160; DB 4; Length 35;

Best Local Similarity 100.0%; Pred. No. 9.9e-13;

Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 36 QARAVGLAGTFRAPFLSRLODLYSIVRRADRAAV 69

Db 1 QARAVGLAGTFRAPFLSRLODLYSIVRRADRAAV 34

RESULT 13

US-09-046-985-7

Sequence 7, Application US/09046985

Patent No. 6121236

GENERAL INFORMATION:

APPLICANT: Ben-Sasson, Shmuel A.

TITLE OF INVENTION: NOVEL MULTIVALENT LIGANDS WHICH MODULATE

TITLE OF INVENTION: ANGIOGENESIS

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: Two Millitia Drive

CITY: Lexington

STATE: Massachusetts

COUNTRY: USA

ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/046,985

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Brook, David E.

REGISTRATION NUMBER: 22,592

REFERENCE/DOCKET NUMBER: CMCC-614

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781) 861-9540

TELEFAX: (781) 861-6240

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 22 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

```

? MOLECULE TYPE: peptide
? FEATURE:
? NAME/KEY: Modified-site
? LOCATION: 1
? OTHER INFORMATION: /label= modified aa
? OTHER INFORMATION: /note= "N-Acetyl-Threonine"
US-09-046-985-7

Query Match      10.4%; Score 101; DB 3; Length 22;
Best Local Similarity 100.0%; Pred. No. 8.4e-06;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      45 TFRAPLSSRLQDLXYSIVRRAD 65
      1 TFRAPLSSRLQDLXYSIVRRAD 21

RESULT 14
US-09-474-743-7
; Sequence 7, Application US/09474743
; Patent No. 6235716
; GENERAL INFORMATION:
; APPLICANT: Ben-Sasson, Shmuel A.
; TITLE OF INVENTION: NOVEL MULTIVALENT LIGANDS WHICH MODULATE
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Millitia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/474,743
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/046,985
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: CMCC-614
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781) 861-6240
; TELEFAX: (781) 861-9540
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1
; OTHER INFORMATION: /label= modified aa
; OTHER INFORMATION: /note= "N-Acetyl-Threonine"
US-09-474-743-7
```

```

Query Match      10.4%; Score 101; DB 4; Length 22;
Best Local Similarity 100.0%; Pred. No. 8.4e-06;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

45 TFRAPLSSRLQDLXYSIVRRAD 65
```

```

Db      1 TFRAPLSSRLQDLXYSIVRRAD 21
      1 TFRAPLSSRLQDLXYSIVRRAD 21

RESULT 15
US-08-740-168A-1
; Sequence 1, Application US/08740168A
; Patent No. 5854205
; GENERAL INFORMATION:
; APPLICANT: O'Reilly, Michael
; APPLICANT: Folkman, M. Judeh
; TITLE OF INVENTION: Therapeutic Antiangiogenic Compositions
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Jones & Askew, LLP
; STREET: 191 Peachtree, 37th Floor
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/740,168A
; FILING DATE: 22-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren, William L.
; REGISTRATION NUMBER: 36,714
; REFERENCE/DOCKET NUMBER: 05213-0223
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: Internal.
; ORIGINAL SOURCE:
; ORGANISM: Murine
; TISSUE TYPE: Collagen
US-08-740-168A-1
```

```

Query Match      9.9%; Score 96; DB 2; Length 20;
Best Local Similarity 85.0%; Pred. No. 3e-05;
Matches 17; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy      1 HSHRDFQPVHLVALNSPLS 20
      1 HTTQDFQPVHLVALNTPLS 20
      1 HSHRDFQPVHLVALNSPLS 20
```

Search completed: August 12, 2002, 10:13:26
Job time: 138 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 10:10:43 ; Search time 30.23 Seconds
(without alignments)
672.395 Million cell updates/sec

Title: US-10-080-797-1
Perfect score: 967
Sequence: 1 HSHRFPQPVYHLVALNSPLS.....SCHHAYIVLCIENSFMWAK 183

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	967	100.0	183	20	AA198693 Human endostatin p
2	967	100.0	183	20	AA192113 SPO ID 76 of MO991
3	967	100.0	183	21	AA1930493 Amino acid sequenc
4	967	100.0	183	21	AA196451 Human endostatin p
5	967	100.0	183	21	AA190771 Human angiogenesis
6	967	100.0	183	21	AA190252 Human angiogenesis
7	967	100.0	183	22	AA190896 Human endostatin(T
8	967	100.0	183	22	AA19379 Human endostatin S
9	967	100.0	216	21	AA190495 Amino acid sequenc
10	967	100.0	684	18	AA196327 Human alpha-1 coll
11	967	100.0	684	20	AA195113 Human alpha1 (XVII

12	967	100.0	1301	20	AA192296 Human alpha-1 (XVI
13	967	100.0	1336	20	AA198694 Human collagen 18
14	963	99.6	183	22	AA194910 Human endostatin p
15	962	99.5	182	21	AA1928399 Human endostatin.
16	962	99.5	182	21	AA194323 Human endostatin p
17	962	99.5	182	21	AA195622 Human endostatin p
18	962	99.5	182	21	AA195622 Human endostatin p
19	962	99.5	182	21	AA195622 Human endostatin p
20	960	99.3	184	22	AA190874 Human endostatin
21	958	99.1	181	22	AA190898 Human endostatin(T
22	956	98.9	271	21	AA1908407 Human endostatin(T
23	954	98.7	180	22	AA190899 Human endostatin(T
24	954	98.7	193	21	AA190877 Human endostatin(T
25	942	97.4	179	22	AA190901 Human endostatin(T
26	937	96.9	178	21	AA194324 Alternate human en
27	937	96.9	178	22	AA190900 Human endostatin(T
28	846	87.5	684	20	AA192514 Mouse alpha1 (XVII
29	845	87.4	184	21	AA190265 Canine angiogenesi
30	840	86.9	184	22	AA1949380 Murine endostatin
31	840	86.9	191	21	AA1949380 Murine endostatin
32	837	86.6	184	20	AA1948409 Endostatin protein
33	837	86.6	184	20	AA1908689 Murine endostatin
34	837	86.6	184	21	AA190258 Murine angiogenesi
35	837	86.6	207	22	AA190231 Murine endostatin
36	837	86.6	207	22	AA190231 Murine endostatin
37	837	86.6	218	20	AA1908691 Murine gene therap
38	837	86.6	580	20	AA1908692 Murine gene therap
39	837	86.6	1288	18	AA1926328 Mouse alpha-1 coll
40	837	86.6	1288	20	AA192297 Mouse alpha-1 (XVI
41	796	82.3	185	20	AA1906197 Anti-angiogenic en
42	771	79.7	184	22	AA1949381 Chicken endostatin
43	544	56.3	180	22	AA1949381 Human endostatin
44	529	54.7	176	21	AA190875 Human endostatin
45	529	54.7	180	22	AA1949382 Human endostatin S

ALIGNMENTS

RESULT 1	
AA198693	AA198693 standard; Protein; 183 AA.
ID	AA198693;
AC	AA198693;
XX	
XX	10-AUG-1999 (first entry)
DE	Human endostatin protein fragment.
KW	Plasminogen; human; angiotatin; endostatin; gene therapy; vector;
KW	anti-angiogenic; attenuation; cytosolic; anti-diabetic; ophthalmology;
KW	tumour growth; solid tumour; diabetic retinopathy; retina.
OS	Homo sapiens.
XX	
PN	WO9926480-A1.
XX	
PD	03-JUN-1999.
XX	
PF	20-NOV-1998; 98WO-US24950.
XX	
PR	20-NOV-1997; 97US-0975424.
XX	
PA	(GENE-) GENETIX PHARM INC.
PA	(MAST) MASSACHUSETTS INST TECHNOLOGY.
XX	
PI	Bachelot T, Leboulch P, Pawluc RJ;
XX	
DR	WPI: 1999-357696/30.
XX	
DR	N-PSDB: AAX77719.
XX	
PT	Anti-angiogenic gene therapy vectors
XX	

PS Disclosure; Page 74-75; 83pp; English.

CC This invention describes a novel viral gene therapy vector comprising a
CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen
CC from human or murine endostatin, human or murine endostatin and
CC angiogenesis-inhibiting fusions and fragments, where the viral vector is
CC sufficiently attenuated for use in human gene therapy. The products of
CC the invention have anti-angiogenic, cytostatic, anti-diabetic and
CC ophthalmological activity. The vector is used in gene therapy for
CC inhibiting tumor growth in humans harbouring a solid tumour. The vector
CC expresses an anti-angiogenic polypeptide. An additional use comprises
CC treatment of diabetic retinopathy, where the anti-angiogenic polypeptide
CC inhibits angiogenesis in the vicinity of the retina. The vector is
CC administered to cells ex vivo and then administered to the patient.

XX Sequence 183 AA:

Query Match 100.0%; Score 967; DB 20; Length 183;
Best Local Similarity 100.0%; Pred. No. 9.8e-110;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HSHRDQPVHLVALNSPLSGMRGIRGADFCQOARAVGLAGTFRATLSRLQDLYSI 60
Db 1 hshrdqpvhlvalnspisgmgirgadfqcqgaravglagtrfatlsrlqdlysl 60
QY 61 VRADRAAVPIVNLKDELFPSSWEALFSGSEGPLKPGARIFSGDKDVLRHPTWPKSW 120
Db 61 vrradraavpivlnlkdelifpswealfsgsegpdkpgarifsfqdkvdlrhpcwpqksv 120
QY 121 HGSDPNGRRLTESYCEWTRREAPSATGQASSLLGRLGQSAASHAHAYVLCIENSFMT 180
Db 121 hgsdpngrlrtesycewtrreapsatgqaassllgrrllgqaasachayivlcienfmc 180
QY 181 ASK 183
Db 181 ask 183

RESULT 2

AAV02113
ID AAV02113 standard; protein; 183 AA.

XX AAV02113;

DT 16-JUL-1999 (first entry)

XX SEQ ID 76 of W09916889.

XX Angiostatin; endostatin; interferon; thrombospondin;
KW interferon-inducible protein; platelet factor 4; anti-angiogenic;
KW anti-tumor; multifunctional protein; angiogenic-mediated disease;
KW cancer; diabetic retinopathy; macular degeneration; arthritis;
KW tumor cell production.

XX Homo sapiens.

OS Homo sapiens.

PN W09916889-A1.

PD 08-APR-1999.

XX 30-SEP-1998; 98WO-US20464.

XX 01-OCT-1997; 97US-0060609.

XX (SEAR) SEARLE & CO G D.

PI Bolanowski MA, Caparon MH, Casperson GF, Gregory SA;

PI Klein BK, Mckearn JP;

XX WPI; 1999-255098/21.

XX New multifunctional proteins useful for treating angiogenic-mediated

PT diseases

XX Disclosure; Page 106-107; 121pp; English.

CC The specification describes multifunctional proteins which comprise
CC combinations of angiostatin, endostatin, interferon, thrombospondin,
CC interferon-inducible protein and platelet factor 4, and have
CC anti-angiogenic and/or anti-tumor activity. The multifunctional protein
CC may exhibit useful properties such as having similar or greater
CC biological activity when compared to a single factor or by having
CC improved half-life or decreased adverse side effects, or a combination
CC of these properties. The proteins can be used for treating an
CC angiogenic-mediated disease, e.g. cancer, diabetic retinopathy, macular
CC degeneration, or arthritis. They can also be used for inhibiting the
CC production of tumor cells (characteristic of lung, breast, ovarian,
CC prostate, pancreatic, gastric, colon, renal, bladder cancers; melanoma,
CC hepatoma, sarcoma and lymphoma) in a patient and for inhibiting tumor
CC growth. The present sequence is used in the course of the invention.

XX Sequence 183 AA:

Query Match 100.0%; Score 967; DB 20; Length 183;
Best Local Similarity 100.0%; Pred. No. 9.8e-110;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HSHRDQPVHLVALNSPLSGMRGIRGADFCQOARAVGLAGTFRATLSRLQDLYSI 60
Db 1 hshrdqpvhlvalnspisgmgirgadfqcqgaravglagtrfatlsrlqdlysl 60
QY 61 VRADRAAVPIVNLKDELFPSSWEALFSGSEGPLKPGARIFSGDKDVLRHPTWPKSW 120
Db 61 vrradraavpivlnlkdelifpswealfsgsegpdkpgarifsfqdkvdlrhpcwpqksv 120
QY 121 HGSDPNGRRLTESYCEWTRREAPSATGQASSLLGRLGQSAASHAHAYVLCIENSFMT 180
Db 121 hgsdpngrlrtesycewtrreapsatgqaassllgrrllgqaasachayivlcienfmc 180
QY 181 ASK 183
Db 181 ask 183

RESULT 3

AAB30493
ID AAB30493 standard; protein; 183 AA.

XX AAB30493;

DT 06-MAR-2001 (first entry)

XX Amino acid sequence of human endostatin encoded by plasmid pMALCH#15.

XX Streptomyces sp. strain C5; SnPA; S. venezuelae; alpha-amylase;

XX endostatin; cancer; tumour growth; angiogenesis.

XX Homo sapiens.

PN W0200060945-A1.

PD 19-OCT-2000.

XX 12-APR-2000; 2000WO-US09747.

XX 13-APR-1999; 99US-0129084.

XX (MERI) MERCK & CO INC.

PI Desanti CL, Strohl WR;

XX WPI; 2000-686970/67.

XX N-PSDB; AAC62023.

PT Preparation of soluble recombinant endostatin involves transforming
 PT Streptomyces host with expression vector comprising nucleotide
 PT sequence encoding endostatin operably linked to linker and leader
 PT peptide -
 XX
 PS Example 1; Fig 6; 57pp; English.
 CC The present sequence represents human endostatin. The protein is
 CC expressed in Streptomyces. Leader sequences of Streptomyces sp. strain
 CC C5 SpA and S. venezuelae alpha-amylase proteins are linked to the
 CC N-terminal of endostatin. This ensures that endostatin protein is
 CC produced as a secreted, soluble protein which needs no refolding, is
 CC stable in the fermentation broth and is produced in large quantities.
 CC The method is used for preparing soluble recombinant human, murine or
 CC primate endostatin, which is useful in the treatment of cancer.
 CC Inhibition of tumour growth, inhibition of angiogenesis, isolation of
 CC receptors for endostatin and for identification of anti-angiogenic
 CC compounds in assays. The endostatin protein is produced as a secreted,
 CC soluble protein which needs no refolding, is stable in the fermentation
 CC broth and is produced in large quantities. Streptomyces are amenable
 CC for cultivation in large fermentations allowing for large quantities of
 CC soluble endostatin to be produced.
 CC
 XX Sequence 183 AA:
 SQ
 Query Match 100.0%; Score 967; DB 21; Length 183;
 Best Local Similarity 100.0%; Pred. No. 9.8e-110;
 Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 HSHRDFQVPLHLVALNSPLSGMGRGIRGADFOQFOARAVGLAGTFRAFLSSRLQDLYSI 60
 Db 1 hshrdfpvhlhvalnslpslsgmrgirgadfcfgqaravglagtfraflssrlqdllysi 60
 QY 61 VRRADRAAVPIVNLKDELLFSPWEALFSGSGEPLKPGARIFSPDGKDYLRHPTWPKSVW 120
 Db 61 vrradraavpivnlkdelifpswealfsgsegplkpgarifsfdgkdvlrhptwpqksvw 120
 QY 121 HGSDPNGRRLTESYCEWTRTEAPSGATGQASSLGGRLGQSAASHAHAYIVLCIENSFMT 180
 Db 121 hgspngrrltesycewtrteapsgatgqassllggrrllgqsaashahayivlcienstf 180
 QY 181 ASK 183
 Db 181 ask 183
 RESULT 4
 AAB16451
 ID AAB16451 standard; Protein; 183 AA.
 AC AAB16451.
 DT 27-OCT-2000 (first entry)
 DE Human endostatin protein sequence.
 DE
 XX Angiogenesis-inhibiting protein receptor; angiogenesis; anglostatin;
 KW endostatin; plasminogen; laminin; treatment; wound healing; solid tumour;
 KW psoriasis; scleroderma; myocardial angiogenesis; Crohn's disease;
 KW cerebral collateral; arteriovenous malformation; rubecosis; cancer;
 KW diabetic retinopathy; arthritis; wound healing; peptic ulcer;
 KW Helicobacter related disease; fracture; cat scratch fever.
 XX
 OS Homo sapiens.
 PN WO200032631-A2.
 PD 08-JUN-2000.
 PF 06-DEC-1999; 99WO-US28897.
 PR 04-DEC-1998; 98US-0206059.

XX
 PA (ENTR-) ENTREMED INC.
 XX
 PI MacDonald NJ, Sim KL;
 XX
 DR WPI; 2000-412290/35.
 XX
 PT New angiogenesis-inhibiting protein receptors, useful in methods for
 PT treating diseases and processes that are mediated by angiogenesis, such
 PT as solid tumours, psoriasis, scleroderma and myocardial angiogenesis -
 XX
 PS Disclosure; Figure 3; 100pp; English.
 CC This invention relates to angiogenesis-inhibiting protein receptors, and
 CC the DNA sequences encoding them. Angiogenesis is the generation of new
 CC blood vessels into a tissue, and normally occurs in wound healing,
 CC foetal and embryonal development and the formation of the corpus luteum,
 CC endometrium and placenta. Anglostatin is a protein (see AAB16450 and
 CC AA68202) involved in angiogenesis, and has an amino acid sequence
 CC similar to that of a plasminogen fragment (see murine plasminogen
 CC AAB16490). Anglostatin has the ability to inhibit angiogenesis.
 CC Endostatin is also an angiogenesis inhibiting protein (see AAB16451 and
 CC AA68203). Sequences AA68242 and AAB16522 represent coding and protein
 CC sequences of human laminin. Laminin is an anglostatin binding protein,
 CC and some of the peptides of the invention share homology with regions of
 CC laminin. Peptides AAB16452-B16521 (excluding AAB16490) are the
 CC angiogenesis-inhibiting protein receptor fragments of the invention. The
 CC peptides bind either anglostatin or endostatin and can be used in methods
 CC for treating diseases and processes that are mediated by angiogenesis,
 CC such as solid tumours, psoriasis, scleroderma, myocardial angiogenesis,
 CC Crohn's disease, cerebral collaterals, arteriovenous malformations,
 CC rubecosis, diabetic retinopathy, arthritis, wound healing, peptic ulcers,
 CC Helicobacter related diseases, fractures, placentaion and cat scratch
 CC fever. They are useful for the detection and prognosis of cancer. DNA
 CC sequences A628204-A628241 encode the peptides of the invention.
 CC
 XX Sequence 183 AA:
 SQ
 Query Match 100.0%; Score 967; DB 21; Length 183;
 Best Local Similarity 100.0%; Pred. No. 9.8e-110;
 Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 HSHRDFQVPLHLVALNSPLSGMGRGIRGADFOQFOARAVGLAGTFRAFLSSRLQDLYSI 60
 Db 1 hshrdfpvhlhvalnslpslsgmrgirgadfcfgqaravglagtfraflssrlqdllysi 60
 QY 61 VRRADRAAVPIVNLKDELLFSPWEALFSGSGEPLKPGARIFSPDGKDYLRHPTWPKSVW 120
 Db 61 vrradraavpivnlkdelifpswealfsgsegplkpgarifsfdgkdvlrhptwpqksvw 120
 QY 121 HGSDPNGRRLTESYCEWTRTEAPSGATGQASSLGGRLGQSAASHAHAYIVLCIENSFMT 180
 Db 121 hgspngrrltesycewtrteapsgatgqassllggrrllgqsaashahayivlcienstf 180
 QY 181 ASK 183
 Db 181 ask 183
 RESULT 5
 AAY90771
 ID AAY90771 standard; Protein; 183 AA.
 AC AAY90771;
 DT 22-AUG-2000 (first entry)
 DE Human angiogenesis inhibiting factor 1 protein.
 DE
 XX Human; angiogenesis inhibiting factor 1; IAF-1; tumour; antibody;
 KW abnormal vessel disease.
 XX

OS Homo sapiens.
 XX CN1244536-A.
 XX 16-FEB-2000.
 PD
 XX 10-AUG-1998; 98CN-0117150.
 PF
 XX 10-AUG-1998; 98CN-0117150.
 PR
 XX 10-AUG-1998; 98CN-0117150.
 XX
 PA (ONCO-) INST ONCOLOGY UNDER TUMOR HOSPITAL CHINE.
 XX
 PI Yang Z, Guo W;
 XX
 DR WPI: 2000-388168/34.
 DR N-PSDB; AAA29884.
 XX
 PT Angiogenesis inhibiting factor 1 and its derivative useful for treating
 PT tumors -
 PS
 XX Claim 1; Fig 5; 41pp; Chinese.
 CC
 CC The present sequence represents an angiogenesis inhibiting factor (1),
 CC designated IAF-1. The present invention also describes: (1) preparation
 CC of (1) and its derivative; (2) an IAF binding acceptor and its
 CC preparation; and (3) an IAF antibody. (1) is useful for preparing new
 CC biological preparations for effectively treating various tumors and
 CC abnormal-vessel diseases. The IAF antibody is preferably a polyclonal
 CC antibody, mosaic antibody, single stranded antibody and human originated
 CC antibody.
 XX
 SQ Sequence 183 AA;
 Query Match 100.0%; Score 967; DB 21; Length 183;
 Best Local Similarity 100.0%; Pred. No. 9.8e-110;
 Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 HSHRDFOPLHLVALNSPLSGMKGIRGADFOCFQARAVGLAGTFRAFLSSRLQDLYST 60
 DB 1 hshrdfgpyhlvalnspisgmgirgadfgcfqaravglagtfraflssrlqdlysl 60
 QY 61 VRRADRAAVPIVNLKDELFPSSWEALFSGSEGLPKGARIFSGDKVLRHPMPQKSW 120
 DB 61 vrradraavpiivnlkdelfpsswealfsgseglpkgarifsgdkvlrhpmpqksvw 120
 QY 121 HGSDPNGRRLTESYCEWTWRAPSATGQASSLLGRLGSAASCHHAYIVLCIENSFMT 180
 DB 121 hgsdpngrrltesycewtwrapsatgqassllgrrlgssaaschhayivlcienfnt 180
 QY 181 ASK 183
 DB 181 ask 183
 RESULT 6
 ID AAY70252 standard; Protein: 183 AA.
 AC AAY70252;
 XX
 DT 06-JUN-2000 (first entry)
 XX
 DE Human angiogenesis inhibitor, endostatin.
 XX
 KW Human; immunoglobulin gamma Fc fragment; endostatin; immunofusion;
 KW angiogenesis; inhibitor; cytoskeletal; antirheumatoid; antiarthritic;
 KW antiproliferative; antidiabetic; ophthalmological; immunosuppressant;
 KW vasculitic; vulnery; treatment; antiarteriosclerosis; tumour;
 KW metastasis; atherosclerosis; psoriasis; rheumatoid arthritis;
 KW ocular angiogenesis disease; diabetic retinopathy; macular degeneration;
 KW myocardial angiogenesis; plaque neovascularisation; telangiectasia;
 KW wound granulation; keloid scar; gene therapy.

XX
 OS Homo sapiens.
 XX
 PN WO200011033-A2.
 XX
 PD 02-MAR-2000.
 XX
 XX 25-AUG-1999; 99WO-US19329.
 PF
 XX 25-AUG-1999; 98US-0097883.
 PR
 XX 25-AUG-1998; 98US-0097883.
 XX
 PA (LEXI-) LEXINGEN PHARM CORP.
 XX
 PI Lo R, Li Y, Gillies SD;
 XX
 DR WPI: 2000-237616/20.
 DR N-PSDB; AA251291.
 XX
 PT Novel fusion protein of angiotensin or endostatin and an immunoglobulin
 PT FC region, useful for treating conditions mediated by angiogenesis,
 PT such as rheumatoid arthritis, tumors and macular degeneration -
 PS
 XX Example 1; Pages 41-42; 68pp; English.
 CC
 CC The patent discloses a DNA molecule encoding a fusion protein comprising
 CC a signal sequence, an immunoglobulin Fc region, and an angiogenesis
 CC inhibitor selected from angiotensin, endostatin, a plasminogen fragment
 CC having angiotensin activity, a collagen XVIII fragment having endostatin
 CC activity, or combinations of them. The fusion protein (immunofusion) is
 CC used to inhibit angiogenesis and to treat diseases or conditions mediated
 CC by angiogenesis. Conditions that may be treated include solid tumours,
 CC blood born tumours, tumour metastasis, benign tumours including
 CC haemangiomas, acoustic neuromas, neurofibromas, trachomas and pyrogenic
 CC granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases
 CC e.g. diabetic retinopathy, retinopathy of prematurity, macular
 CC degeneration, corneal graft rejection, neovascular glaucoma, retrolental
 CC fibroplasia, rubeosis and Osler-Webber syndrome, myocardial angiogenesis,
 CC plaque neovascularisation, telangiectasia, haemophilic joints,
 CC angiodysplasia, wound granulation, and excessive or abnormal stimulation of
 CC endothelial cells, intestinal cells, atherosclerosis, scleroderma and
 CC hypertrophic scars, i.e. keloid scars. The DNA constructs may be used
 CC in gene therapy. The present sequence is a human endostatin used in the
 CC construction of immunofusion containing human immunoglobulin gamma
 CC (19c) Fc fragment.
 XX
 SQ Sequence 183 AA;
 Query Match 100.0%; Score 967; DB 21; Length 183;
 Best Local Similarity 100.0%; Pred. No. 9.8e-110;
 Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 HSHRDFOPLHLVALNSPLSGMKGIRGADFOCFQARAVGLAGTFRAFLSSRLQDLYST 60
 DB 1 hshrdfgpyhlvalnspisgmgirgadfgcfqaravglagtfraflssrlqdlysl 60
 QY 61 VRRADRAAVPIVNLKDELFPSSWEALFSGSEGLPKGARIFSGDKVLRHPMPQKSW 120
 DB 61 vrradraavpiivnlkdelfpsswealfsgseglpkgarifsgdkvlrhpmpqksvw 120
 QY 121 HGSDPNGRRLTESYCEWTWRAPSATGQASSLLGRLGSAASCHHAYIVLCIENSFMT 180
 DB 121 hgsdpngrrltesycewtwrapsatgqassllgrrlgssaaschhayivlcienfnt 180
 QY 181 ASK 183
 DB 181 ask 183
 RESULT 7
 ID AAU00896 standard; Protein: 183 AA.
 XX

AAU00896;
04-JUL-2001 (first entry)
Human Endostatin(TM) protein.
Human; Endostatin(TM); angiogenesis mediated disease; solid tumours;
blood borne tumour; leukaemia; tumour metastasis; benign tumour;
haemangioma; acoustic neuroma; neurofibroma; trachoma; rubecosis;
pyogenic granuloma; rheumatoid arthritis; psoriasis; colon cancer;
ocular angiogenic disease; diabetic retinopathy; macular degeneration;
retinopathy of prematurity; macular corneal rejection;
neovascular glaucoma; retrolental fibroplasia; Osler-Weber Syndrome;
myocardial angiogenesis; plaque neovascularisation; telangiectasia;
haemophilic joint; angiodiroma; wound granulation.
Homo sapiens.
WO200119989-A2.
22-MAR-2001.
14-SEP-2000; 2000WO-US25166.
14-SEP-1999; 99US-0153698.
(ENTR-) ENTREMED INC.
Liang H, Sim KL, Chang-Murad A, Zhou X, Madsen J, Boerner RJ,
Bermejo LL, Mistry FR, Shepard SR, Schrimsher JL;
WPI; 2001-244802/25.
N-PSDB; AAS00867.
Producing Endostatin protein for treating angiogenesis mediated
diseases such as solid tumours, comprises recombinantly producing the
protein using an expression system, and recovering and purifying the
protein -
Claim 5; Page 29; 67pp; English.
The sequence represents Human Endostatin(TM). The new method of the
invention is useful for producing, recovering and purifying Endostatin
(TM) from biological sources, such as biological fluids, tissues, cells,
culture media, and fermentation media. Endostatin(TM) is useful for
treating angiogenesis mediated diseases such as solid tumours, blood
borne tumours, leukaemias, tumour metastases, benign tumours, e.g.
haemangioma, acoustic neuromas, neurofibromas, trachomas, and pyogenic
granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases,
e.g., diabetic retinopathy, retinopathy of prematurity, macular
degeneration, corneal graft rejection, neovascular glaucoma, colon
cancer, retrolental fibroplasia, rubecosis, Osler-Weber Syndrome,
myocardial angiogenesis, plaque neovascularisation, telangiectasia,
haemophilic joints, angiodiroma, and wound granulation. Endostatin(TM)
is also useful for treating disease of excessive or abnormal stimulation
of endothelial cells such as intestinal adhesions, atherosclerosis,
scleroderma and hypertrophic scars. Higher yields of more purified, and
biologically active Endostatin(TM) are obtained by the new method.
Endostatin(TM) can be stored in buffers for extended periods of time, and
also subjected to lyophilisation, while preserving biological activity.
Centrifugation of broth from fermentation steps in production is avoided,
preventing unwanted potential cellular lysis and contamination with
additional proteins, pigments, enzymes and other cellular chemicals and
debris.

Sequence 183 AA;

Query Match 100.0%; Score 967; DB 22; Length 183;

Best Local Similarity 100.0%; Pred. No. 9,8e-110; Indels 0; Gaps 0;
Matches 183; Conservative 0; Mismatches 0;

QY 1 HSHRDFPVHLVALNSPLSGMGRGADGFCQFOARAVGLACTFRAFLSSRLQDIYSI 60

|||||
Db 1 hshrdfpvhlvalnspisgmgirgadfqcfcgqaravglactfiraflssrlqdiysi 60
QY 61 VRRADRAAVPIVNLKDELLFPSEALFSGSGPLKPGARIRFSFGKDVLRHPTPQKSW 120
|||||
Db 61 vrradraavpivnlkdelllfswaalfsgsgplkpgarirfsfdgkdvlrhptwpqksvw 120
QY 121 HGSDPNGRRLTESYCEMTRTEAPSATGQASSLLGRRLLGQSAASCHHAYIYLCIENSFMT 180
|||||
Db 121 hgsdpngrlrtesycemtrteapsatgqassllgrrllgqsaaschnhayivlcienstmc 180
QY 181 ASK 183
|||
Db 181 ask 183

RESULT 8

AAB49379
ID AAB49379 standard; Protein; 183 AA.
XX

AC AAB49379;
XX

DT 02-MAR-2001 (first entry)
XX

DE Human endostatin SEQ ID NO: 2.
XX

KW Endostatin; antiangiogenic; angiogenesis; human; mouse; chicken;
KW cancer; inflammation; angiogenesis-dependent disease.
OS Homo sapiens.

XX WO200067771-A1.
XX

PN 16-NOV-2000.
XX

PD 02-MAY-2000; 2000WO-US12063.
XX

PP 06-MAY-1999; 99US-0132907.
XX

PR 14-JUL-1999; 99US-0353333.
XX

PS (BURN-) BURNHAM INST.
XX

PI Vuori K;
XX

XX WPI; 2001-040937/05.
XX

DR N-PSDB; AAC88289.
XX

XX Endostatin peptide comprising at least four endostatin amino acid
PT residues are e.g. angiogenesis inhibitors for treating cancer and
PR diabetic retinopathy -

XX Disclosure; Fig 1; 146pp; English.
PS

XX The present invention provides endostatin peptides which can be used in
XX the modulation of angiogenesis. This is useful in the treatment of
CC cancers, inflammation, rheumatoid arthritis, chronic articular

CC rheumatism, psoriasis, disorders associated with inopportune invasion of
CC vessels such as diabetic retinopathy, neovascular glaucoma, retinopathy

CC of prematurity, macular degeneration, corneal graft rejection,
CC retrolental fibroplasia, rubecosis, capillary proliferation in

CC atherosclerotic plaques and osteoporosis. Other angiogenesis-dependent
CC diseases include Osler-Weber syndrome, myocardial angiogenesis, plaque

CC neovascularisation, telangiectasia, haemophilic joints and wound
CC granulation. In addition, the peptides can be used as birth control

CC agents.
CC

XX Sequence 183 AA;

QY 1 HSHRDFPVHLVALNSPLSGMGRGADGFCQFOARAVGLACTFRAFLSSRLQDIYSI 60

Query Match 100.0%; Score 967; DB 22; Length 183;

Best Local Similarity 100.0%; Pred. No. 9,8e-110; Indels 0; Gaps 0;
Matches 183; Conservative 0; Mismatches 0;

OY 1 HSHRDQPVHLVALNSPLSGMRGIRGADFOCFQOARAVGLAGTRAPLSSRLDLYST 60
DB 1 hshrdqpvhlvalnspisgmrgirgadfgcqqaravglagtrfalsrldlyst 60
OY 61 VRADRAAVPIVNLKDELLFPSSWEALFSGSEGLPKPGARIFSDGKDVLRHPTWPKSW 120
DB 61 vrradraavpivnlkdelifpswealifsgseglpkpgarilfsdgkdvlrhptwpskw 120
OY 121 HGSDPNGRRLTESYCEWTWREAPSATGOASSLLGGRLLGOSASCHHAYIVLCIENSFMT 180
DB 121 hgsdpngrrltesyccetwreapatsatgqassllgrrllggsaaschhayivlcienfnt 180
OY 181 ASK 183
DB 181 ask 183
RESULT 9
AAB30495
ID AAB30495 standard; protein: 216 AA.
XX AAB30495;
XX 06-MAR-2001 (first entry)
DT
XX Amino acid sequence of vaa-endostatin fusion protein in pANT3052.
DE
XX Streptomyces sp. strain C5; SnPA; S. venezuelae; alpha-amy-lase;
KM endostatin; cancer; tumour growth; angiogenesis.
XX
OS Synthetic.
OS Streptomyces sp.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..28 /note="vaa signal sequence"
FT Protein 29..216
FT /note="endostatin"
XX
XX MO200060945-A1.
XX
XX 19-OCT-2000.
XX
XX 12-APR-2000; 2000MO-US09747.
XX
XX 13-APR-1999; 99US-0129084.
XX
XX (MERI) MERCK & CO INC.
XX
XX Desanti CL, Strohl WR;
PI
XX
XX WPI: 2000-686970/67.
DR N-PSDB; AAC62025.
XX
XX Preparation of soluble recombinant endostatin involves transforming
PT Streptomyces host with expression vector comprising nucleotide
PT sequence encoding endostatin operably linked to linker and leader
PT peptide
XX
XX Example 1; Fig 10A-B; 57pp; English.
XX
XX The present sequence represents a fusion protein of vaa and endostatin.
XX The specification describes a method for the production of soluble,
XX recombinant human endostatin in Streptomyces. Leader sequences of
XX Streptomyces sp. strain C5 SnPA and S. venezuelae alpha-amy-lase proteins
XX are linked to the N-terminal of endostatin. This ensures that endostatin
XX protein is produced as a secreted, soluble protein which needs no
XX refolding, is stable in the fermentation broth and is produced in large
XX quantities. The method is used for preparing soluble recombinant human,
XX murine or primate endostatin, which is useful in the treatment of cancer,
XX inhibition of tumour growth, inhibition of angiogenesis, isolation of
XX receptors for endostatin and for identification of anti-angiogenic

CC compounds in assays. The endostatin protein is produced as a secreted,
CC soluble protein which needs no refolding, is stable in the fermentation
CC broth and is produced in large quantities. Streptomyces are amenable
CC for cultivation in large fermentations allowing for large quantities of
CC soluble endostatin to be produced.
XX
XX Sequence 216 AA:
SQ
OY 1 HSHRDQPVHLVALNSPLSGMRGIRGADFOCFQOARAVGLAGTRAPLSSRLDLYST 60
DB 34 hshrdqpvhlvalnspisgmrgirgadfgcqqaravglagtrfalsrldlyst 93
OY 61 VRADRAAVPIVNLKDELLFPSSWEALFSGSEGLPKPGARIFSDGKDVLRHPTWPKSW 120
DB 94 vrradraavpivnlkdelifpswealifsgseglpkpgarilfsdgkdvlrhptwpskw 153
OY 121 HGSDPNGRRLTESYCEWTWREAPSATGOASSLLGGRLLGOSASCHHAYIVLCIENSFMT 180
DB 154 hgsdpngrrltesyccetwreapatsatgqassllgrrllggsaaschhayivlcienfnt 213
OY 181 ASK 183
DB 214 ask 216
RESULT 10
AAM26327
ID AAM26327 standard; protein: 684 AA.
XX
XX AAM26327;
AC
XX 19-NOV-1997 (first entry)
DT
XX Human alpha-1 collagen (XVIII).
XX
XX Alpha-1 collagen; type XVIII collagen; cartilage degeneration.
XX
XX Homo sapiens.
XX
XX
XX Key Location/Qualifiers
FH Peptide 1..6 /label="GXGX'Y'__motif"
FT /note="Claim 1"
FT
FT Peptide 7..12 /label="GXGX'Y'__motif"
FT /note="Claim 1"
FT
FT Peptide 13..18 /label="GXGX'Y'__motif"
FT /note="Claim 1"
FT
FT Peptide 19..24 /label="GXGX'Y'__motif"
FT /note="Claim 1"
FT
FT Peptide 25..30 /label="GXGX'Y'__motif"
FT /note="GXGX'Y'__motif"
FT
FT Peptide 31..36 /label="GXGX'Y'__motif"
FT /note="Claim 1"
FT
FT Peptide 37..42 /label="GXGX'Y'__motif"
FT /note="GXGX'Y'__motif"
FT
FT Peptide 48..53 /label="GXGX'Y'__motif"
FT /note="Claim 1"
FT
FT Peptide 54..59 /label="GXGX'Y'__motif"
FT /note="GXGX'Y'__motif"
FT
FT Peptide 74..79 /label="Claim 1"

FT	/label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	80..85 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	86..91 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	92..97 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	98..103 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	104..109 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	110..115 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	121..126 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	/label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	129..134 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	135..140 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	141..146 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	147..152 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	153..158 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	159..164 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	165..170 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	171..176 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	/label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	181..186 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	187..192 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	193..198 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	215..220 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	221..226 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	227..232 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	233..238 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	239..244 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	257..262 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
<hr/>	
PT	Peptide
FT	/note= "Claim 1" 263..268 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	269..274 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	275..280 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	286..291 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	292..297 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	309..314 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	315..320 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	322..328 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	329..334 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	335..340 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	354..359 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	360..365 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	366..372 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	523..528 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	542..547 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	590..595 /label= GXYGX'Y'_motif /note= "Claim 1"
FT	Peptide
FT	US5643783-A.
PX	01-JUL-1997.
PD	01-DEC-1993; 93US-0159784.
PX	01-DEC-1993; 93US-0159784.
PR	(HARD) HARVARD COLLEGE.
PX	Oh SP, Olsen BR;
PI	WPI; 1997-350247/32.
XX	N-PDSB; AAT84484.
DR	
XX	Nucleic acid encoding human alpha-1 collagen - for production of recombinant alpha-1 collagen, for use in the treatment of cartilage degeneration
PT	Claim 1: Column 23-30; 35PD: Enq1ish.
PS	

XX Novel human type alpha-1 (XVIII) collagen is characterized by
CC 10 triple helical domains containing the GXYGX'Y' motif (where X,
CC Y, X' and Y' represent any amino acid), the helical domains being
CC separated and flanked by non-triple helical regions which may
CC provide flexibility. Alpha-1 collagen is expressed in multiple
CC tissues, especially liver, lung and kidney. A claimed plasmid
CC comprising alpha-1 collagen nucleic acid (see AAT84484) and an
CC expression control sequence can be used to express recombinant
CC collagen in prokaryotic or eukaryotic (especially mammalian) host
CC cells. The alpha-1 collagen may be used to treat a patient
CC suffering from a disease associated with cartilage degradation, and
CC for supplementing collagen. It can also be used as a connective

Query Match 100.0%; Score 967; DB 18; Length 684;
Best Local Similarity 100.0%; Pred. No. 7e-109;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HSHRDPOVLHVALNSPLSGMRGIRGADFCQOARAVGLACTPRAFLSSRLDLYSI 60
DB 502 hshrdtpvhlvalnspisgmrgirgadfcqgaravglactfratfissrlqdllysl 561
OY 61 VRRADRAAVPIVNLKDELPPSWMALFSGSEGPLKPGARIFSPDGKDVLRHPMPQKSW 120
DB 562 vrradraavpivnlkdelppswmalfsgeegplkpgariftfdgkdvlrhpcwpqksw 621
OY 121 HGSDPNGRRLTESYCETWRTAPSATGOASSLLGGRLLGOSASCHHAYIVLCIENSFMT 180
DB 622 hgsdpgngrrltesycetwrtapsgatgassllggrllggsaaschaylvlcienfnt 681
OY 181 ASK 183
DB 682 ask 684

RESULT 11

AA25113
ID AAY25113 standard; Protein; 684 AA.
XX
AC AAY25113;
XX
DT 25-AUG-1999 (first entry)
XX
DE Human alpha1 (XVIII) collagen protein.
XX

KW Alpha1(XVIII) collagen; mimetic; endostatin; atomic coordinate; library;
KW anti-angiogenic; heparin binding domain; receptor binding domain; mimic;
KW alpha-helix A domain; carbohydrate recognition domain; CRD domain;
KW treatment; angiogenesis; tumour; human.
XX

OS Homo sapiens.
XX
PN WO9931616-A1.
XX
PD 24-JUN-1999.
XX
PF 16-DEC-1998; 98WO-US26783.
XX
PR 16-DEC-1997; 97US-0069727.
XX
PA (HARD) HARVARD COLLEGE.
XX

PI Hohenester E, Olsen BR, Sasaki T, Timpl R;
XX
DR WPI; 1999-395243/33.
DR N-PSDB; AAX78379.
XX
PT Identifying mimetics of mammalian endostatin
XX
PS Disclosure; Fig 5A-C; 75pp; English.
XX

CC This invention describes a novel method for identifying mimetics of
CC mammalian endostatin. The method comprises identifying a compound

CC having atomic coordinates with non-trivial similarity to selected
CC coordinates of atoms of a mammalian endostatin involves (a) providing
CC a library of atomic coordinates of compounds in a library of candidate
CC compounds, (b) comparing the library of atomic coordinates to the
CC selected coordinates of a mammalian endostatin and (c) selecting from the
CC library at least one candidate compound on the basis of selection
CC criteria which include similarities between the atomic coordinates of the
CC selected candidate compound and the atomic coordinates of the mammalian
CC endostatin. The invention also describes the use of an anti-angiogenic
CC fragment of endostatin comprising a domain selected from a heparin
CC binding domain, a receptor binding domain, and exposed on alpha-helix A
CC domain, and a carbohydrate recognition domain (CRD) domain. The methods
CC can be used for designing and selecting endostatin mimics. The compounds
CC identified can be used for treating undesired angiogenesis, e.g. tumours.
CC This sequence represents human alpha1(XVIII) collagen which is used in
CC the description of the method.
XX

SO Sequence 684 AA;

Query Match 100.0%; Score 967; DB 20; Length 684;
Best Local Similarity 100.0%; Pred. No. 7e-109;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HSHRDPOVLHVALNSPLSGMRGIRGADFCQOARAVGLACTPRAFLSSRLDLYSI 60
DB 502 hshrdtpvhlvalnspisgmrgirgadfcqgaravglactfratfissrlqdllysl 561
OY 61 VRRADRAAVPIVNLKDELPPSWMALFSGSEGPLKPGARIFSPDGKDVLRHPMPQKSW 120
DB 562 vrradraavpivnlkdelppswmalfsgeegplkpgariftfdgkdvlrhpcwpqksw 621
OY 121 HGSDPNGRRLTESYCETWRTAPSATGOASSLLGGRLLGOSASCHHAYIVLCIENSFMT 180
DB 622 hgsdpgngrrltesycetwrtapsgatgassllggrllggsaaschaylvlcienfnt 681
OY 181 ASK 183
DB 682 ask 684

RESULT 12

AAW92296
ID AAW92296 standard; peptide; 1301 AA.
XX

AC AAW92296;
XX
DT 28-APR-1999 (first entry)
XX

DE Human alpha-1 (XVIII) collagen chain common sequence HUI8(common)36.
XX
KW Human; type XVIII collagen; liver disease; cirrhosis; detection;
KW hepatocellular carcinoma; diagnosis.
XX

OS Homo sapiens.
XX
PN WO9856399-A1.
XX
PD 17-DEC-1998.
XX
PF 12-JUN-1998; 98WO-US12327.
XX
PR 12-JUN-1997; 97US-0049369.
XX

PA (EFT-) ACAD FINLAND.
PA (FIBR-) FIBROGEN INC.
PA (INRM) INST NAT SANTE & RECH MEDICALE.
XX

PI Clement B, Pihlajaniemi T, Rehn M;
XX
DR WPI; 1999-070292/06.
XX

PT Diagnosis and monitoring of liver disease by measuring collagen type

PT XVIII levels - with elevated levels indicative of disease,
PT especially cirrhosis or hepatocellular carcinoma
XX
PS Example 6; Fig 8; 56pp; English.
XX
CC A method has been developed for the detecting liver disease. The method
CC comprises: (a) reacting a patient sample with antibodies (Ab) specific
CC for collagen type XVIII (Coll18); (b) measuring the amount of Ab-antigen
CC complex (C) formed as indicator of the amount of Coll18 present; (c)
CC similar analysis of a non-diseased control; and (d) comparing the
CC amounts of Coll18 in the two samples to detect presence or progression of
CC disease. Elevated levels of Coll18 are: (i) indicative of disease,
CC specifically cirrhosis; and (ii) predictive of the prognosis of disease,
CC specifically hepatocellular carcinoma (there is a relationship between
CC Coll18 mRNA levels and tumour size and necrosis, and survival times are
CC significantly higher in patients with higher Coll18 levels). The method
CC provides non-invasive, early and accurate diagnosis of liver disease.
CC The present sequence represents the sequence common to human alpha-1
CC (XVIII) collagen chain from the present invention.
XX
SQ Sequence 1301 AA:

Query Match 100.0%; Score 967; DB 20; Length 1301;
Best Local Similarity 100.0%; Pred. No. 1.8e-108;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HSHRDPOVHLVALNPSLSGMRGIRGADFCQOARAVGLAGTFPAFLSSRIQDIXSI 60
Db 1119 hshrdtpqvhlvalnpslsgmrgirgadfcqgaravglaqtfafissrlqdlxsi 1178
QY 61 VRRADRAAVPIVNLKDELFPSEWALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVW 120
Db 1179 vrradraavpivnlkdelfpsewalfsgeglkpgarifsfqgkvlnhptwpqksvw 1238
QY 121 HGSDPNGRRILTESYCEFTWRTAPSATGQASSLLGRRLLGQSAASCHHAYIVLCIENSFMT 180
Db 1239 hgsdpngrriltesycefwtreapatsatgqassllgrrllggsaaschhayivlcienfnt 1298
QY 181 ASK 183
Db 1299 ask 1301

RESULT 13
AA08694
ID AAY08694 standard; Protein: 1336 AA.
XX
AC AAY08694;
XX
DT 10-AUG-1999 (first entry)
XX
DE Human collagen 18 protein.
XX
KW Plasminogen; human; angiotatin; endostatin; gene therapy; vector;
KW anti-angiogenic; attenuation; cytostatic; anti-diabetic; ophthalmology;
KW tumour growth; solid tumour; diabetic retinopathy; retina; collagen 18.
XX
OS Homo sapiens.
XX
PN WO926480-A1.
XX
PD 03-JUN-1999.
XX
PF 20-NOV-1998; 98WO-US24950.
XX
PR 20-NOV-1997; 97US-0975424.
XX
PA (GENE-) GENETIX PHARM INC.
XX (MASS) MASSACHUSETTS INST TECHNOLOGY.
XX
PI Bachelot T, Leboulch P, Pawliuk RJ;
XX

DR WPI; 1999-357696/30.
DR N-PSDB; AAX77720.
XX
PT Anti-angiogenic gene therapy vectors
XX
PS Disclosure; Page 77-80; 83pp; English.
XX
CC This invention describes a novel viral gene therapy vector comprising a
CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen
CC from human or murine angiotatin, human or murine endostatin and
CC angiogenesis-inhibiting fusions and fragments, where the viral vector is
CC sufficiently attenuated for use in human gene therapy. The products of
CC the invention have anti-angiogenic, cytostatic, anti-diabetic and
CC ophthalmological activity. The vector is used in gene therapy for
CC inhibiting tumour growth in humans harbouring a solid tumour. The vector
CC expresses an anti-angiogenic polypeptide. An additional use comprises
CC treatment of diabetic retinopathy, where the anti-angiogenic polypeptide
CC inhibits angiogenesis in the vicinity of the retina. The vector is
CC administered to cells ex vivo and then administered to the patient.
XX
SQ Sequence 1336 AA:

Query Match 100.0%; Score 967; DB 20; Length 1336;
Best Local Similarity 100.0%; Pred. No. 1.9e-108;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HSHRDPOVHLVALNPSLSGMRGIRGADFCQOARAVGLAGTFPAFLSSRIQDIXSI 60
Db 1154 hshrdtpqvhlvalnpslsgmrgirgadfcqgaravglaqtfafissrlqdlxsi 1213
QY 61 VRRADRAAVPIVNLKDELFPSEWALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVW 120
Db 1214 vrradraavpivnlkdelfpsewalfsgeglkpgarifsfqgkvlnhptwpqksvw 1273
QY 121 HGSDPNGRRILTESYCEFTWRTAPSATGQASSLLGRRLLGQSAASCHHAYIVLCIENSFMT 180
Db 1274 hgsdpngrriltesycefwtreapatsatgqassllgrrllggsaaschhayivlcienfnt 1333
QY 181 ASK 183
Db 1334 ask 1336

RESULT 14
AAB49810
ID AAB49810 standard; Protein: 183 AA.
XX
AC AAB49810;
XX
DT 02-MAR-2001 (first entry)
XX
DE Human endostatin peptide fragment SEQ ID NO: 23.
XX
KW Endostatin; antiangiogenic; angiogenesis; human; mouse; chicken;
KW cancer; inflammation; angiogenesis-dependent disease.
XX
OS Homo sapiens.
XX
PN WO200067771-A1.
XX
PD 16-NOV-2000.
XX
PF 02-MAY-2000; 2000WO-US12063.
XX
PR 06-MAY-1999; 99US-0132907.
XX PR 14-JUL-1999; 99US-0353333.
XX
PA (BURN-) BURHAM INST.
XX
PI Vuori K;
XX
DR WPI; 2001-040937/05.

XX Endostatin peptide comprising at least four endostatin amino acid
PT residues are e.g. angiogenesis inhibitors for treating cancer and
XX 'diabetic retinopathy'.
PS Claim 17; Page 125-126; 146pp; English.
XX
CC The present invention provides endostatin peptides which can be used in
CC the modulation of angiogenesis. This is useful in the treatment of
CC cancers, inflammation, rheumatoid arthritis, chronic articular
CC rheumatism, psoriasis, disorders associated with inappropriate invasion of
CC vessels such as diabetic retinopathy, neovascular glaucoma, retinopathy
CC of prematurity, macular degeneration, corneal graft rejection,
CC retrolental fibroplasia, rubeosis, capillary proliferation in
CC atherosclerotic plaques and osteoporosis. Other angiogenesis-dependent
CC diseases include Osler-Webber syndrome, myocardial angiogenesis, plaque
CC neovascularisation, telangiectasia, haemophilic joints and wound
CC granulation. In addition, the peptides can be used as birth control
CC agents.
XX
SQ Sequence 183 AA:

Query Match 99.6%; Score 963; DB 22; Length 183;
Best Local Similarity 99.5%; Pred. No. 3e-109;
Matches 182; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HSHRDFQPVHLVALNSPLSGMGRIGRADFOCFQOARAVAGLAGTFRATLSSRLQDLYSI 60
DB 1 hshrdtqpvhlvalnspisgmgrigradfcfgqaravaglagtfratfslsqdlysl 60
OY 61 VRRADRAAVPIVNLKDELFPSSWEALFSGSEGPLKPGARIFSDGKDVLRHPTWPKSW 120
DB 61 vrradraavpivnlkdelifpswealfsgseglpkparifsfdgkdvlrhptwpqksv 120
OY 121 HGSDPNGRRLTESYCEWTREAPSATGQASLLGRLGQSAASHAYIVLCIENSFWT 180
DB 121 hgsdpngrlrtesyctwrtreapscatqgassllgtrllgqsaaschayivlcienfwt 180
OY 181 ASK 183
DB 181 ask 183

RESULT 15
AAB28399
ID AAB28399 standard; Protein; 182 AA.
XX
AC AAB28399;
XX
DT 19-FEB-2001 (first entry)
XX
DE Human endostatin.
XX
KW Human; endostatin; cytosolic; antiproliferative;
KW vascular endothelial growth factor; VEGF; antibody; VEGF2 receptor;
KW cancer; vascularised solid tumour.
XX
OS Homo sapiens.
XX
PN WO200064946-A2.
XX
PD 02-NOV-2000.
XX
PF 28-APR-2000; 2000WO-US11367.
XX
PR 28-APR-1999; 99US-0131432.
XX
PA (TEXA) UNITV TEXAS SYSTEM.
XX
PI Thorpe PE, Brekken RA;
XX
UR WPI; 2000-687317/67.

XX Immunogenic composition for the treatment and diagnosis of cancer
PT comprises an anti-VEGF (vascular endothelial growth factor) antibody
XX binding the same epitope as the monoclonal antibody ATCC PTA 1595 -
XX
PS Example 10; Page 291-292; 298pp; English.
XX
XX

CC The present invention relates to anti-Vascular Endothelial Growth Factor
CC (VEGF) antibodies that bind to the same epitope as the monoclonal
CC antibody ATCC PTA 1595 and which significantly inhibit VEGF binding to
CC the VEGF receptor VEGFR2, without inhibiting VEGF binding to the VEGF
CC receptor VEGFR1. The present sequence is human endostatin. Endostatin
CC may be conjugated onto the anti-VEGF antibodies of the present invention.
CC The anti-VEGF antibodies of the present invention are useful for the
CC treatment and diagnosis of cancer, especially vascularised solid tumours.
XX
SQ Sequence 182 AA:

Query Match 99.5%; Score 962; DB 21; Length 182;
Best Local Similarity 100.0%; Pred. No. 4e-109;
Matches 182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HSHRDFQPVHLVALNSPLSGMGRIGRADFOCFQOARAVAGLAGTFRATLSSRLQDLYSI 60
DB 1 hshrdtqpvhlvalnspisgmgrigradfcfgqaravaglagtfratfslsqdlysl 60
OY 61 VRRADRAAVPIVNLKDELFPSSWEALFSGSEGPLKPGARIFSDGKDVLRHPTWPKSW 120
DB 61 vrradraavpivnlkdelifpswealfsgseglpkparifsfdgkdvlrhptwpqksv 120
OY 121 HGSDPNGRRLTESYCEWTREAPSATGQASLLGRLGQSAASHAYIVLCIENSFWT 180
DB 121 hgsdpngrlrtesyctwrtreapscatqgassllgtrllgqsaaschayivlcienfwt 180
OY 181 AS 182
DB 181 as 182

Search completed: August 12, 2002, 10:13:06
Job time: 143 sec